The Annals of Statistics
1974, Vol. 2, No. 1, 21-38

IMPROVING ON EQUIVARIANT ESTIMATORS[1]

BY J. F. BREWSTER AND J. V. ZIDEK[2]

University of Manitoba and University of British Columbia

Techniques for improving on equivariant estimators are described. They may be applied, although without assurance of success, whatever be the family of underlying distributions. The loss function is required to satisfy an intuitively reasonable condition but is otherwise arbitrary. One of these techniques amounts to a sample space, orbit-by-orbit analysis of the conditional expected loss given the orbit. It yields, when successful, a "testimator". A second technique obtains the limit of a certain sequence of "testimator-like" estimators. The result is "smoother" than a testimator and often identical to a generalized Bayes estimator over much of its domain. Applications are presented. In the first we extend results of Stein (1964) and obtain a minimax estimator which is generalized Bayes, and in a univariate subcase, admissible within the class of scale-equivariant estimators. In the second, we extend a result of Srivastava and Bancroft (1967).

1. Introduction. This paper describes techniques for determining for a given equivariant estimator, $\phi^{(\circ)}$, another whose risk function is never larger than that of $\phi^{(\cdot)}$. The loss function must satisfy a natural requirement but may otherwise be arbitrary. We place no restriction on the underlying family. For some families and loss functions technical difficulties may make application of these techniques unfeasible, and, in other cases, the estimator obtained by applying one of these techniques may be identical to the original estimator! So our methods cannot be offered with a guarantee that they will always succeed. But they have been successful in many cases including those discussed in Section 2, and they do provide simple proofs and generalizations of results already obtained by other authors.

Two closely related methods are discussed in this paper. The first involves conditioning on an appropriately chosen statistic: the second involves taking the limit of a sequence of "testimators" (a term possibly coined by Sclove) involving tests based on this statistic. In order to illustrate these techniques consider independent observations $X_1, X_2, \cdots, X_m$ from a normal population with unknown mean μ and unknown variance σ^2. Letting $S^2 = \sum (X_i - \bar{X})^2$ and looking only at estimators of the form cS^2, it is easy to see that $(m+1)^{-1}S^2$ is the "best" estimator of σ^2 under squared error loss.

Received August 1972; revised February 1973.

[1] This research was supported by a grant from the National Research Council of Canada.

[2] The second author is on leave at the University College, London.

AMS 1970 *subject classifications.* Primary 62C15; Secondary 62F10.

Key words and phrases. Equivariant estimators, normal variance estimators, maximean estimators, minimax generalized Bayes estimators, improving estimators, bowl-shaped loss functions.

But now consider estimators of the form $\phi(Z)S^2$, where $Z = m^{\frac{1}{2}}|\bar{X}|/S$. And for each μ, σ, let $\phi_{\mu,\sigma}$ represent the best choice of ϕ. Since $E_{\mu,\sigma}[(\phi(Z)S^2 - \sigma^2)^2] = E_{\mu,\sigma}[E_{\mu,\sigma}[(\phi(Z)S^2 - \sigma^2)^2 \mid Z]]$, $\phi_{\mu,\sigma}(z)$ will be that value of c which minimizes $E_{\mu,\sigma}[(cS^2 - \sigma^2)^2 \mid Z = z]$. In other words, $\phi_{\mu,\sigma}(z) = E_{\mu,\sigma}[S^2 \mid Z = z]/E_{\mu,\sigma}[S^4 \mid Z = z]$.

In order to compare $\{\phi_{\mu,\sigma}\}$ with $(m+1)^{-1}$, we first notice that $\phi_{\mu,\sigma}(z) \leqq \phi_{0,1}(z) = (m+2)^{-1}(1+z^2)$, for all z, μ, σ. Moreover, for $0 \leqq z < (m+1)^{-\frac{1}{2}}$, $\phi_{0,1}(z) < (m+1)^{-1}$. And since $E_{\mu,\sigma}[(cS^2 - \sigma^2)^2 \mid Z = z]$ is a "strictly bowl-shaped" (in fact, convex) function of c, an improvement will therefore be obtained, for all μ and σ, if $(m+1)^{-1}$ is replaced by $\phi_{0,1}(z)$, whenever $0 \leqq z < (m+1)^{-\frac{1}{2}}$. The resulting estimator

$$\min\{(m+1)^{-1}, (m+2)^{-1}(1+Z^2)\}S^2, \tag{1.1}$$

is precisely Stein's estimator (1964).

Let us look at this problem again, but consider a smaller class of estimators. To this end, fix $r > 0$, and consider only estimators of the form $\phi(Z)S^2$, where $\phi(z) \equiv c$, $0 \leqq z \leqq r$, and $\phi(z) \equiv d$, $z > r$. For such an estimator, $E_{\mu,\sigma}[(\phi(Z)S^2 - \sigma^2)^2] = E_{\mu,\sigma}[(cS^2 - \sigma^2)^2 \mid Z \leqq r] \cdot P_{\mu,\sigma}[Z \leqq r] + E_{\mu,\sigma}[(dS^2 - \sigma^2)^2 \mid Z > r] \cdot P_{\mu,\sigma}[Z > r]$ and it follows that $c_{\mu,\sigma}(r) = E_{\mu,\sigma}[S^2 \mid Z \leqq r]/E_{\mu,\sigma}[S^4 \mid Z \leqq r]$ and $d_{\mu,\sigma}(r) = E_{\mu,\sigma}[S^2 \mid Z > r]/E_{\mu,\sigma}[S^4 \mid Z > r]$ are the best choices for the constants, for given μ, σ. Here $c_{\mu,\sigma}(r) \leqq c_{0,1}(r) < (m+1)^{-1}$, and therefore, letting $\phi_r(z) = c_{0,1}(r)$, $0 \leqq z \leqq r$, and $\phi_r(z) = (m+1)^{-1}$, $z > r$, $\phi_r(Z)S^2$ has uniformly smaller risk than $(m+1)^{-1}S^2$. This is precisely Brown's estimator (1968).

But now select $0 < r' < r$. By repeating the previous argument and noticing that $c_{0,1}(r') < c_{0,1}(r)$, we are able to conclude that $\phi_{r',r}(Z)S^2$ has uniformly smaller risk than $\phi_r(Z)S^2$, where $\phi_{r',r}(z) = c_{0,1}(r')$, $0 \leqq z \leqq r'$; $\phi_{r',r}(z) = c_{0,1}(r)$, $r' < z \leqq r$ and $\phi_{r',r}(z) = (m+1)^{-1}$, $z > r$. We can clearly continue to produce step-function (inadmissible) estimators by selecting successively smaller constants. But notice that the starting point, r, is arbitrary.

Now, for each $i = 1, 2, 3, \cdots$, select a finite partition of $[0, \infty)$, represented by $0 = r_{i0} < r_{i1} < \cdots < r_{in_i} = \infty$, and a corresponding estimator $\phi^{(i)}(Z)S^2$, where $\phi^{(i)}(z) = E_{0,1}(S^2 \mid Z \leqq r_{ij})/E_{0,1}(S^4 \mid Z \leqq r_{ij})$, $r_{i,j-1} < z \leqq r_{ij}$. Then, providing $\max_{1 \leqq j \leqq n_i - 1} |r_{ij} - r_{i,j-1}| \to 0$ and $r_{i,n_i-1} \to \infty$, the sequence $\phi^{(i)}$ will converge pointwise to ϕ^{**}, where

$$\phi^{**}(z) = E_{0,1}(S^2 \mid Z \leqq z)/E_{0,1}(S^4 \mid Z \leqq z). \tag{1.2}$$

Although we are, in general, unable to compare $\phi^{(i)}$ and $\phi^{(i')}$, $i \neq i'$, we do know that $\phi^{(i)}$ is better than $(m+1)^{-1}$, for all i. It would not be surprising, therefore, if the same were true of ϕ^{**}, and we shall show in Section 2 that this is the case. In fact, $\phi^{**}(Z)S^2$ is also generalized Bayes and admissible within the class of scale-equivariant procedures.

The preceding paragraphs, culminating in Stein's estimator and $\phi^{**}(Z)S^2$, illustrate the two techniques to be discussed in this paper. Until now, invariance has not been mentioned as it is not essential in describing the methods, although it certainly plays an important role in the previous examples. However, the description below is limited to equivariant estimators, because we have not yet found a broader context in which the methods can be more precisely formulated and realized. Nevertheless, there are many potential applications of this work, especially since maximum likelihood estimators are often equivariant.

In a problem which remains invariant under a group $\mathscr{G}$ acting transitively on the parameter space, there is an obvious method for improving on a given $\mathscr{G}$-equivariant estimator. For in this case the risk function of any $\mathscr{G}$-equivariant estimator is constant. So there is a "best" $\mathscr{G}$-equivariant estimator and it is better than, if different from, the given estimator.

In some cases, even the best $\mathscr{G}$-equivariant estimator may be easily improved. If there is a proper subgroup $\mathscr{K} \subset \mathscr{G}$, which also acts transitively on Θ, there is a best $\mathscr{K}$-equivariant estimator, and it is better than, if different from, the best $\mathscr{G}$-equivariant estimator. An example wherein the best equivariant estimator is improved on in this manner is given in James and Stein (1961). Their example concerns the estimation of the multivariate normal covariance matrix.

Although our methods apply in the situations described above, they are suggested because of their possible value in other situations where improvement cannot be so simply achieved. They apply when $\bar{\mathscr{G}}$ is exactly transitive on Θ and even when $\bar{\mathscr{G}}$ is not transitive.

Incidentally, since the best $\mathscr{G}$-equivariant estimator, $\phi^{(\circ)}$, is usually a generalized Bayes procedure, our methods indirectly offer a means of assessing certain improper priors. These priors are obtained from the right Haar measure on $\mathscr{G}$ (see, for example, Zidek (1969)) and are often thought of as expressing prior ignorance. But this choice of prior is somewhat arbitrary. And if another estimator, ϕ, of uniformly smaller risk than $\phi^{(\circ)}$ is found, then since any "proper" Bayesian would prefer ϕ to $\phi^{(\circ)}$, some doubt is cast on the suitability, in this case, of right Haar measure as a prior. The method given in Theorem 3.2 has even produced in some cases (see Section 2.2) a ϕ, superior to $\phi^{(\circ)}$, which is also generalized Bayes.

Much has been written on the subject to which this paper is devoted (the most relevant references being cited in the remarks of Section 2). Undoubtedly, the most interesting example where improvement on the best fully equivariant estimator is possible is contained in the work of Stein on the estimation of the multivariate normal mean in dimensions greater than 2 (see, for example, James and Stein (1961)). Our methods formally apply in this situation but they are unsuccessful, yielding only the original estimator and not a superior alternative, like Stein's. As we show below, our techniques rely, essentially, on orbit-by-orbit reductions in the expected loss and, as the work of Sclove, Morris and Radhakrishnan (1972) suggest, any such attempt to improve on the usual estimator is destined to fail.

Our methods apply also to interval estimation (Brewster (1972)). In fact, the notion of "recognizable" or "relevant" subsets which arose in the context of interval estimation (see, for example, Buehler (1959)) is clearly involved in our methods.

The two techniques in this paper appear first (see Section 2) in exemplary applications. Then they are described (see Section 3) in a fairly general setting where we reveal features of examples which have enabled them to be treated by these techniques. These examples include, in addition to those in Section 2, others given in Brewster (1972) and Zidek (1971, 1973).

Our techniques require that a subgroup, $\mathscr{H} \subset \mathscr{G}$, be selected; the search for an estimator better than a given $\mathscr{G}$-equivariant (or, in some cases, $\mathscr{H}$-equivariant) estimator is restricted to a class of $\mathscr{H}$-equivariant estimators. No insight is offered on the matter of choosing $\mathscr{H}$ (see Stein (1964) for some remarks in this connection). We found a suitable $\mathscr{H}$ quite easily in applications where a choice existed.

The second of the two methods is given in Theorem 3.2 and does provide, as heuristics suggest it might, a smoother alternative to $\phi^{(\cdot)}$ than does the first method. It is technically more complex than the first method and is not so generally applicable. But where it has been applicable, dramatic results have sometimes emerged. For example, for each of a variety of loss functions, it produces (see Section 2.1) a minimax, generalized Bayes estimator for the variance of the normal law with unknown mean which is different from the usual estimator.

2. Applications. In each application considered in this section the loss function is specified in terms of a nonnegative function, $W(u)$, whose domain D, say, is either $(0, \infty)$ or $(-\infty, \infty)$ and whose minimum is attained at u_0, where $u_0 = 1$ or 0 according as $D = (0, \infty)$ or $(-\infty, \infty)$. It will always be assumed that $W(u)$ is continuous and strictly bowl-shaped, that is, strictly decreasing for $u \leqq u_0$ and strictly increasing for $u \geqq u_0$. As a consequence W is differentiable almost everywhere, and we will assume, whenever necessary for integrals involving W, that interchange of integral and derivative is permissible. Presumably W need not be continuous and our results could be proved in a more general setting using the methods of Brown (1968) involving generalized derivatives, but we have not fully explored this possibility.

We will say that a family of density functions, $\mathscr{F} = \{f(x \mid \tau): x \in \mathscr{X} \subset (-\infty, \infty), \tau \in \mathscr{T} \subset (-\infty, \infty)\}$, has the monotone likelihood ratio property (MLRP) if $x_1 < x_2$ and $\tau_1 < \tau_2$ implies $f(x_1 \mid \tau_2)/f(x_1 \mid \tau_1) < f(x_2 \mid \tau_2)/f(x_2 \mid \tau_1)$. For such a family it is well known (see, for example, Lehman (1959)) that $\tau \to \int h(x)f(x \mid \tau)\,dx$ has at most one sign change if h has one sign change. We use this property throughout the following lemma whose proof is immediate.

LEMMA 2.1. *If f is a density on* $(0, \infty)$ $[(-\infty, \infty)]$ *and* $\{f(xc^{-1}): c > 0\}$ $[\{f(x - c): -\infty < c < \infty\}]$ *has the* MLRP, *then*

$$c \to \int xW'(cx)f(x)\,dx \qquad \left[\int W'(x + c)f(x)\,dx\right]$$

has at most one sign changeand $c \to \int W(cx)f(x)\,dx$ $[\int W(x+c)f(x)\,dx]$ *is strictly bowl-shaped (or monotone).*

2.1. *Estimating the normal variance with unknown mean.* Here we adopt the canonical form of the general linear model and suppose $X \sim MVN_p(\mu, \tau I)$ and $U \sim MVN_n(0, \tau I)$ are to be independently observed. On the basis of these observations, τ is to be estimated, where loss is given by $L(\hat{\tau}; \mu, \tau) = W(\hat{\tau}\tau^{-1})$. A sufficient statistic in this problem is (X, T), where if $||\cdot||$ denotes the usual Euclidean norm, $T = ||U||^2$. We consider only nonrandomized estimators which are a function of this statistic.

The problem remains invariant under the transformation group $\mathscr{G}$ under which

$$(X, T) \to (\alpha\Gamma X + \beta, \alpha^2 T)$$
$$(2.1.1) \qquad (\mu, \tau) \to (\alpha\Gamma\mu + \beta, \alpha^2\tau)$$
$$\hat{\tau} \to \alpha^2\hat{\tau}$$

where $\alpha > 0$, $\beta \in R^p$ and Γ is a $p \times p$ orthogonal matrix. It follows that any nonrandomized $\mathscr{G}$-equivariant estimator of τ is of the form cT, for some constant $c > 0$. Since $\mathscr{G}$ acts transitively on the parameter space, the risk function of cT,

$$E_{\mu,\tau} W(cT\tau^{-1}) = E_{0,1} W(cT)\,,$$

is independent of the unknown parameters, and we assume that $E_{0,1}W(cT)$ is not a monotone function of c. Then there is an optimum choice for c, say $c^{(\circ)}$, whose existence is established in the next lemma. The proof is an immediate consequence of Lemma 2.1.

LEMMA 2.1.1. *The function,* $c \to E_{0,1} W(cT)$, *is strictly bowl-shaped and uniquely minimized at* $c = c^{(\circ)}$ *satisfying*

$$(2.1.2) \qquad E_{0,1} W'(c^{(\circ)}T)T = 0\,.$$

Thus $c^{(\circ)}T$ is the best $\mathscr{G}$-equivariant estimator.

Let $\mathscr{H}$ denote the subgroup of $\mathscr{G}$ obtained by requiring in (2.1.1) that $\beta = 0$ and that Γ be a diagonal orthogonal matrix. Any $\mathscr{H}$-equivariant estimator is of the form $\phi(Z)T$ where $Z = (Z_1, \cdots, Z_p)'$ and $Z_i = |X_i|T^{-\frac{1}{2}}$, $i = 1, \cdots, p$. The risk of such an estimator is

$$\begin{aligned} r(\phi; \mu, \tau) &= E_{\mu,\tau} W[\phi(Z)T\tau^{-1}] \\ &= E_{\zeta,1} W[\phi(Z)T] \\ &= r(\phi; \zeta) \end{aligned}$$

say, where $\zeta = (\zeta_1, \cdots, \zeta_p)'$ and $\zeta_i = |\mu_i|\tau^{-\frac{1}{2}}$, $i = 1, \cdots, p$. Since we deal here only with $\mathscr{H}$-equivariant estimators we may assume without loss of generality that $\tau = 1$.

We represent X_i^2 by a chi-squared random variable with $1 + 2K_i$ degrees of freedom, where K_i denotes a Poisson random variable with mean $\lambda_i = \frac{1}{2}\zeta_i^2$, and the K_i, $i = 1, \cdots, p$, are independent of each other and of T. Let $K = (K_1, \cdots, K_p)$, and observe that the joint density of T and Z conditional on $K = k = (k_1, \cdots, k_p)$ is

$$f(t, z \mid k) \propto t^{\frac{1}{2}(n+p)+k_{\bullet}-1} \exp[-\tfrac{1}{2}t(1 + \|z\|^2)] \prod_{i=1}^{p} z_i^{2k_i} ,$$

independent of ζ, where $k_{\bullet} = \sum k_i$.

For the model under consideration, we now obtain results analogous to those described in Section 1 for estimating the normal variance under quadratic loss.

As a simple consequence of Lemma 2.1, we have the following result.

LEMMA 2.1.2. *The function, $c \to E[W(cT) \mid Z = z, K = k]$ is strictly bowl-shaped and uniquely minimized at $c = \phi_k(z)$ satisfying*

$$E(W'[\phi_K(Z)T]T \mid Z = z, K = k) = 0 .$$

For any estimator $\phi(Z)T$ define ϕ^* by

$$\phi^*(z) = \min \{\phi(z), \phi_0(z)\} .$$

THEOREM 2.1.1. *For any ζ*

$$r(\phi^*; \zeta) \leqq r(\phi; \zeta)$$

with inequality if $P_\zeta[\phi^*(Z) \neq \phi(Z)] > 0$.

PROOF. Observe that

$$r(\phi; \zeta) = E_\zeta E(W[\phi(Z)T] \mid Z, K) .$$

Denote the inner, conditional expectation evaluated at $Z = z$ and $K = k$ by $r[\phi(z) \mid z, k]$. Either $\phi^*(z) = \phi(z)$, when $r[\phi^*(z) \mid z, k] = r[\phi(z) \mid z, k]$, or $\phi^*(z) = \phi_0(z) < \phi(z)$. In the latter case since $c \to r[c \mid z, k]$ is strictly bowl-shaped and, as is easily shown using Lemma 2.1, $\phi_k(z) \leqq \phi_0(z)$ for all k, it follows that $r[\phi^*(z) \mid z, k] < r[\phi(z) \mid z, k]$. The conclusion of the theorem now follows.

COROLLARY 2.1.1. *If $\phi^{(\circ)}(z) \equiv c^{(\circ)}$, then for any ζ,*

$$r(\phi^{(\circ)*}; \zeta) < r(\phi^{(\circ)}; \zeta) .$$

PROOF. Let $r[c \mid z, k]$ be defined as in the proof of Theorem 2.1.1. Then

$$r[c \mid z, 0] \propto \int W(ct) t^{\frac{1}{2}(n+p)-1} \exp[-\tfrac{1}{2}t(1 + \|z\|^2)]\, dt \propto EW(c'T)T^{\frac{1}{2}p} ,$$

where $c' = c(1 + \|z\|^2)^{-1}$. Since $r[c \mid z, 0]$ is strictly bowl-shaped in c, so is $EW(c'T)T^{\frac{1}{2}p}$ in c' and its minimum is attained at $c' = \phi_0(z)(1 + \|z\|^2)^{-1} = c^*$, say. But $EW'(c^{(\circ)}T)T \cdot T^{\frac{1}{2}p} > (1/c^{(\circ)})^{\frac{1}{2}p} \cdot EW'(c^{(\circ)}T)T = 0$ (see Equation (2.1.2)). So $\phi_0(z)(1 + \|z\|^2)^{-1} < c^{(\circ)}$ and $\phi_0(Z) < c^{(\circ)}$ with positive probability for each ζ. The conclusion now follows from Theorem 2.1.1.

Note that with c^* defined as in the proof of Corollary 2.1.1, $\phi_0(z) = c^*(1 + \|z\|^2)$ with c^* a constant and $c^* < c^{(\circ)}$. Moreover,

$$\phi^{(\circ)*}(Z)T = \min \{c^{(\circ)}, c^*(1 + \|Z\|^2)\}T .$$

EXAMPLE 2.1.1. Let $W_1(x) = |x - 1|$. If U_α denotes the random variable with density $\propto u^{\frac{1}{2}\alpha}e^{-\frac{1}{2}u}$, $u \geqq 0$, and median (U_α) denotes its median value, then it is easily seen that c^* and $c^{(\circ)}$ are given by

$$c^* = [\text{median}\,(U_{n+p})]^{-1}\,, \qquad c^{(\circ)} = [\text{median}\,(U_n)]^{-1}\,.$$

EXAMPLE 2.1.2. Let $W_2(x) = (x - 1)^2$. Then

$$c^* = (n + p + 2)^{-1}\,, \qquad c^{(\circ)} = (n + 2)^{-1}\,.$$

EXAMPLE 2.1.3. Brown (1968) has shown that the usual unbiased estimator is obtained when $W(x) = x - 1 - \ln x = W_u(x)$, say. For this case, where under-estimation as well as over-estimation is heavily penalized,

$$c^* = (n + p)^{-1}\,, \qquad c^{(\circ)} = n^{-1}\,.$$

EXAMPLE 2.1.4. Brown also looks at $W(x) = (\ln x)^2 = W_{\ln}(x)$, say. Here, under-estimation is rather more heavily penalized than over-estimation and, using the notation introduced in Example 2.1.1,

$$c^* = \exp[-E \ln U_{n+p-2}]\,, \qquad c^{(\circ)} = \exp[-E \ln U_{n-2}]\,.$$

Note that $(n + p)^{-1} < c^* < (n + p - 2)^{-1}$, while $n^{-1} < c^{(\circ)} < (n - 2)^{-1}$. Comparing these results with those obtained in the last two examples reveals the expected result that the size of the estimate increases with the cost of under-estimation.

We now present another method for improving on $c^{(\circ)}T$. We select the class of sets, $\mathscr{C} = \{C(r) : 0 < r \leqq \infty\}$, where $C(r) = \{z \in [0, \infty)^p : 0 \leqq ||z||^2 \leqq r\}$.

LEMMA 2.1.3. *The function, $c \to E[W(cT) \mid Z \in C, K = k]$ is, for every k, $C \in \mathscr{C}$, strictly bowl-shaped and uniquely minimized at $c = \psi_k(C)$ satisfying*

$$E[W'(\psi_K(C)T)T \mid Z \in C, K = k] = 0\,.$$

PROOF. $E[W(cT) \mid Z \in C, K = k] \propto \int W(ct)h_k(t; C)\,dt$ where $h_k(t; C) = \int_C f(t, z \mid k)\,dz$. For simplicity let $h(t) = h_k(t; C)$. According to Lemma 2.1, the required result will be obtained when we show that $\{h(t\tau^{-1}) : \tau > 0\}$ has the MLRP. But

$$h(\theta t) \propto t^{\frac{1}{2}n-1}e^{-\frac{1}{2}\theta t} \int_{\theta C} [\exp(-\tfrac{1}{2}t||z||^2)] \prod z_i^{k_i - \frac{1}{2}}\,dz$$

and if $\theta_1 < \theta_2$, $h(\theta_1 t)h^{-1}(\theta_2 t)$ must be shown to be increasing in t. This is easily accomplished by differentiation and by exploiting the monotonicity of $\theta \to (\int_{\theta C} ||z||^2 g(z)\,dz) \cdot (\int_{\theta C} g(z)\,dz)^{-1}$ for any positive function, g.

The following result is an application of Lemma 2.1.

LEMMA 2.1.4. *$\psi_k(C) \leqq \psi_0(C)$ for all k, and if C is a proper subset of D, $\psi_0(C) < \psi_0(D)$.*

Note that $\psi_0[C(\infty)] = c^{(\circ)}$, so $\psi_0(C) < c^{(\circ)}$ for each $C \in \mathscr{C}$ with $C \neq C(\infty)$.

LEMMA 2.1.5. *Suppose $\phi(Z)T$ is such that $\phi(z) \equiv \alpha > \psi_0(C)$, $z \in C$. Define*

$\phi_C(z)$ as $\phi(z)$ or $\psi_0(C)$ according as $z \notin C$ or $z \in C$. Then $\phi_C(Z)T$ has everywhere smaller risk function than $\phi(Z)T$.

PROOF.

$$r(\phi; \zeta) = E_\zeta[E(W[\phi(Z)T] \mid Z \in C, K) \cdot P_\zeta(Z \in C \mid K) + E(W[\phi(Z)T] \mid Z \notin C, K) \cdot P_\zeta(Z \notin C \mid K)]$$

and the first of the two inner expectations evaluated at $K = k$ is, since $\psi_k(C) \leqq \psi_0(C) < \alpha$,

$$E(W[\alpha T] \mid Z \in C, K = k) > E(W[\psi_0(C)T] \mid Z \in C, K = k)$$

by the bowl-shaped property established in Lemma 2.1.3.

The following theorem is a simple consequence of Lemmas 2.1.4 and 2.1.5.

THEOREM 2.1.2. *Let $C_2 \subset C_1 \subset C(\infty)$, $C_i \in \mathscr{C}$, with both inclusions being proper, $\phi^{(0)}(z) \equiv c^{(0)}$, and $\phi^{(i)}(z) = \phi^{(i-1)}(z)$ or $\psi_0(C_i)$ according as $z \notin C_i$ or $z \in C_i$, $i = 1, 2$. Then $r(\phi^{(2)}; \zeta) < r(\phi^{(1)}; \zeta) < r(\phi^{(0)}; \zeta)$, for all ζ.*

Define ϕ^{**} by

$$\phi^{**}(z) = \psi_0[C(||z||^2)] \,. \tag{2.1.3}$$

We shall show that the risk function of $\phi^{**}(Z)T$ is never larger than that of $c^{(0)}T$, and also that, in a sense which will be made precise, $\phi^{**}(Z)T$ is generalized Bayes. For the loss functions introduced in Examples 2.1.1–2.1.4, ϕ^{**} is easily computed. If V is the positive random variable whose joint density with Z is $\propto v f_{T,Z}(v, z \mid \zeta)$, then the respective ϕ^{**}'s are found to be $\text{median}_{\zeta=0}(V^{-1} \mid ||Z|| \leqq ||z||)$, $E_{\zeta=0}(T \mid ||Z|| \leqq ||z||)/E_{\zeta=0}(T^2 \mid ||Z|| \leqq ||z||)$, $E_{\zeta=0}^{-1}(T \mid ||Z|| \leqq ||z||)$, and $\exp\{-E_{\zeta=0}(\ln T \mid ||Z|| \leqq ||z||)\}$. The condition $K = 0$ is replaced by $\zeta = 0$ because the joint density of T and Z is the same under either condition.

LEMMA 2.1.6. *$\phi^{**}(z)$ is strictly increasing in $||z||^2$ and continuous.*

PROOF. The monotonicity is a direct consequence of Lemma 2.1.4. Fix $r_0 > 0$, and suppose $r_n \to r_0$, $r_n > r_{n+1}$. Notice that $\psi_0[C(r_n)]$ is a strictly decreasing sequence, which therefore has a limit $\alpha \geqq \psi_0[C(r_0)]$. From the bowl-shaped property established in Lemma 2.1.3, it follows that

$$\int W'(\alpha t) t f(t \mid Z \in C(r_n), K = 0)\, dt < 0 \,.$$

So, if the interchange of limit and integral is justified,

$$\int W'(\alpha t) t f(t \mid Z \in C(r_0), K = 0)\, dt \leqq 0 \,.$$

It follows that $\alpha \leqq \psi_0[C(r_0)]$, that is, $\alpha = \psi_0[C(r_0)]$. The required interchange is justified by the dominated convergence theorem. This requires the regularity assumptions on W which entail the existence of a function $H(t) \geqq |W'(\alpha t) \cdot t|$, which is integrable with respect to the conditional density, $f(t \mid Z \in C(r_1), K = 0)$.

Also required is the inequality, $f(t \mid Z \in C(r_n), K = 0)/f(t \mid Z \in C(r_1), K = 0) \leqq P(\|Z\|^2 \leqq r_1 \mid K = 0)/P(\|Z\|^2 \leqq r_0 \mid K = 0)$. A similar argument can be used to demonstrate the continuity from below.

THEOREM 2.1.3. *For any* ζ,

$$r(\phi^{**}; \zeta) \leqq r(\phi^{(\circ)}; \zeta) .$$

PROOF. Let $\{r_{ij} : i = 1, 2, \cdots; j = 1, 2, \cdots, n_i\}$ be any double sequence for which

(a) $0 = r_{i_0} < r_{i_1} < \cdots < r_{in_i} = \infty$,
(b) $\{r_{ij} : j = 1, 2, \cdots, n_i\} \subset \{r_{i+1,j} : j = 1, 2, \cdots, n_{i+1}\}$,
(c) $r_{in_i-1} \to \infty$ as $i \to \infty$, and
(d) $\lim_{i\to\infty} \max_{1\leq j\leq n_i-1} |r_{ij} - r_{i,j-1}| = 0$.

Define

$$\phi^{(i)}(z) = \psi_0[C(r_{ij})] , \qquad z \in C(r_{ij}) \sim C(r_{i,j-1}) .$$

Proceeding inductively from Theorem 2.1.2, $r(\phi^{(i)}; \zeta) < r(\phi^{(\circ)}; \zeta)$, for all ζ. But

$$\lim_{i\to\infty} \phi^{(i)}(z) = \phi^{**}(z) ,$$

for all z, by the continuity of ϕ^{**}. On applying Fatou's lemma, the desired conclusion is obtained.

Now let $\mathscr{S}$ denote the scale subgroup of $\mathscr{G}$ obtained by requiring in (2.1.1) that $\beta = 0$, and that Γ be the identity matrix. Any $\mathscr{S}$-equivariant estimator of τ is of the form $\delta(Y)T$ where $Y = XT^{-\frac{1}{2}}$. The risk function for such an estimator, $r(\delta; \eta)$, depends on μ, τ only through $\eta = \mu\tau^{-\frac{1}{2}}$. Thus, if a (possibly improper) prior measure is given on the range of η, R^p, the (possibly generalized) Bayes procedure among $\mathscr{S}$-equivariant estimators may be determined. Such a procedure is obtained by setting $\delta(Y) = \phi^{**}(Z)$.

THEOREM 2.1.4. *Within the class of $\mathscr{S}$-equivariant estimators, $\phi^{**}(Z)T$ is generalized Bayes with respect to the prior on $\eta \in R^p$ with density*

$$\pi(\eta) = \int_0^\infty \nu^{p/2-1}(1 + \nu)^{-1} \exp(-\nu\|\eta\|^2/2)\, d\nu . \tag{2.1.4}$$

PROOF. Let Π be any σ-finite measure on R^p and $g_\Pi(t, y) = \int_{R^p} f_{T,Y}(t, y \mid \eta)\, d\Pi(\eta)$. If, for each y, $\psi_\Pi(y)$ minimizes and makes finite

$$\int W(ct) g_\Pi(t, y)\, dt$$

as a function of c, then $\psi_\Pi(Y)T$ is the Bayes or generalized Bayes procedure with respect to Π among $\mathscr{S}$-equivariant procedures.

If we ignore factors depending on y, we are led to search for a Π which satisfies

$$g_\Pi(t, y) \propto f_{T \mid \|Y\| \leq \|y\|}(t \mid 0) \propto t^{\frac{1}{2}(n+p)-1} e^{-\frac{1}{2}t} \int_0^{\|y\|^2} w^{\frac{1}{2}p-1} e^{-\frac{1}{2}tw}\, dw .$$

But

$$g_\Pi(t, y) \propto t^{\frac{1}{2}(n+p)-1} e^{-\frac{1}{2}t} \int_{R^p} \exp(-\tfrac{1}{2}\|t^{\frac{1}{2}}y - \eta\|^2)\, d\Pi(\eta) .$$

Thus on setting $d\xi(\eta) = e^{-\frac{1}{2}||\eta||^2} d\Pi(\eta)$ and $x = t^{\frac{1}{2}}y$, we obtain

$$\int_{R^p} e^{\eta' x} d\xi(\eta) \propto \int_0^1 (1 - v)^{\frac{1}{2}p-1} e^{\frac{1}{2}||x||^2 v} dv .$$

Observe that the quantity on the right is a mixture of moment generating functions for normal distributions. This observation readily leads to the choice for ξ and hence Π.

REMARK. It follows that $\phi^{**}(Z)T$ is also generalized Bayes (see, for example, Zidek (1969)).

THEOREM 2.1.5. *Suppose* $p = 1$, *and* $W(u) = (u - 1)^2$. *Then* $\phi^{**}(Z)T$ *is admissible within the class of* $\mathscr{S}$*-equivariant estimators.*

PROOF. It is easily shown that for π as in Theorem 2.1.4,

$$\int_1^\infty [\pi(\eta)\eta^2]^{-1} d\eta = \int_{-\infty}^{-1} [\pi(\eta)\eta^2]^{-1} d\eta = \infty .$$

The theorem is now an immediate consequence of a result of Brewster (1972).

REMARKS. It may well be unnecessary to introduce the Poisson variables $K_1, K_2, \cdots, K_p$ as we have done. In fact, all of our analysis can be done by conditioning only on Z if the loss function is convex or one of those introduced in Examples 2.1.1–2.1.4. For general bowl-shaped loss functions we are prevented from proving our main results by conditioning on Z alone, only because of our failure, due to technical difficulties, to obtain the necessary analogues of Lemmas 2.1.2 and 2.1.3. The proofs would follow the lines of the argument suggested in the introduction and since the joint densities of T and Z given $K = 0$ and given $\zeta = 0$ are identical, the same estimators would be obtained. The introduction of the auxiliary Poisson variables is objectionable because it means that we have exploited fortuitous special features of the normal law and in doing so have departed from the general methods described in Section 3 which call for conditioning on Z alone.

The choice of $\mathscr{C}$ is somewhat arbitrary. We could as well have taken $\mathscr{C} = \{C(r) : 0 < r \leqq \infty\}$ where $C(r) = \{z \in [0, \infty)^p : 0 \leqq z_i \leqq r, i = 1, \cdots, p\}$ and obtained a result analogous to Theorem 2.1.3. But our results for varying $\mathscr{C}$ are too incomplete for inclusion here.

We could have treated without additional difficulty the problem of estimating τ^q, q being a known constant, but have declined to do so to avoid the notational clutter which would have ensued.

The estimator, say $\hat{\tau}$, obtained from Theorem 2.1.1 when quadratic loss is assumed, is obtained by Stein (1964). Stein confines his search for $\hat{\tau}$ to the class of scale—orthogonal—equivariant estimators, that is, those of the form $F(||Z||^2)T$ for some function F; here $\hat{\tau}$ emerges from the larger class of estimators of the form $\phi(Z)T$, although we also could have introduced orthogonality from the outset.

To prove the superiority of $\hat{\tau}$ over the usual estimator, Stein represents $||Z||^2$ in terms of a central chi-squared random variable with $p + 2L$ degrees of freedom for a certain Poisson variable L, and then conditions on L alone; in our proof (for quadratic loss, at least) we condition on Z alone.

Inspiration for our work is also found in that of Brown (1968). Brown introduces, in the context of estimation, the notion of a bowl-shaped loss function and reveals the role played by the MLRP, when loss is bowl-shaped. His work applies here formally, in the case $p = 1$, and yields a result similar to that of Lemma 2.1.5. This is not the best possible result as our Theorem 2.1.2 goes on to show, but, as Brown suggests, his goal was not to obtain the best possible result for particular families but rather, to show that for a wide class of families, the best affine-equivariant estimator is inadmissible.

Recently, Strawderman (1972) obtained, when $p = 1$ and loss is squared error, a class of estimators which are superior to the usual one. These estimators are of the form $\phi(Z)T$ where ϕ is required to satisfy certain properties, and Stein's estimator (1964) is a member of this class. Strawderman shows that, for certain ε and a, the generalized Bayes rule corresponding to the prior measure on (ζ, τ) with density given by

$$\pi(\zeta, \tau)\, d\tau\, d\zeta \propto \tau^{-\frac{1}{2}\varepsilon} \int_0^1 v^{\frac{1}{2}-a}(1 - v)^{-\frac{1}{2}} \exp[-\tfrac{1}{2}\zeta^2 v(1 - v)^{-1}]\, dv(d\tau/\tau)\, d\zeta\,,$$

is a member of this class. It can be shown that if $\varepsilon = 0$ and $a = 1$, then the estimator obtained is that given by equation (2.1.3) when loss is quadratic.

Looking, for simplicity, only at quadratic loss, it is perhaps instructive to compare Stein's estimator, $\phi^{(\circ)*}(Z)T$, and $\phi^{**}(Z)T$. Notice first that $\phi^{**}(0) = \phi^{(\circ)*}(0)$, and for all $z \neq 0$, $\phi^{**}(z) < \phi^{(\circ)*}(z)$. Therefore, looking at the best estimator when $\zeta = 0$, it is clear that $r(\phi^{(\circ)*}, 0) < r(\phi^{**}, 0)$. Of course, in view of Theorem 2.1.5 and the scale-inadmissibility of Stein's estimator, there exists $\zeta_0 \neq 0$ such that $r(\phi^{**}, \zeta_0) < r(\phi^{(\circ)*}, \zeta_0)$, at least when $p = 1$.

Now, it is evident that the origin is not important in the problem. Therefore, if ξ is an arbitrary vector in R^p, then an estimator possessing properties similar to $\phi^{**}(Z)T$ is given by $\phi_\xi^{**}(X, T)T$, where

$$\phi_\xi^{**}(X, T) = \phi^{**}(T^{-1}||X - \xi||^2)\,.$$

Moreover, $\phi_\xi^{**}(X, T)T$ has an obvious interpretation. First notice that the natural (minimax and admissible) estimator for τ in the analogous problem with known mean ξ is given by

$$\phi(X, T; \xi) = (n + p + 2)^{-1}(T + ||X - \xi||^2)\,.$$

In the problem with unknown mean μ, we now let ξ represent a preliminary estimate of μ, and use estimator $\phi_\xi^{**}(X, T)T$ to estimate τ. If we find, in fact, that $X \equiv \xi$ (supporting our prior suspicion), then our estimate will agree with that given by $\phi(X, T; \xi)$. Otherwise, the estimate is modified, depending on the (normalized) distance, $T^{-\frac{1}{2}}||X - \xi||$, between X and ξ. As the distance becomes infinite the estimate approaches $(n + 2)^{-1}T$.

This interpretation is similar to that of Stein's estimator,

$$\phi^{(\circ)*}(T^{-1}||X - \xi||^2)T = \min\{\phi(X, T; \xi), (n + 2)^{-1}T\}\,.$$

This estimator may thus be regarded as the result of first testing the hypothesis $\mu = \xi$ at a particular significance level—using $\phi(X, T; \xi)$ if the hypothesis is accepted, and $(n + 2)^{-1}T$ if the hypothesis is rejected.

Finally, it is interesting to observe that the method of proof used to construct ϕ^{**} enables us to demonstrate the minimaxity of a number of other estimators. That is, if $\phi(z)$ is a non-decreasing function of z, and $\phi^{**}(z) \leqq \phi(z) \leqq (n + 2)^{-1}$, for all z, then $\phi(Z)T$ is a minimax estimator of τ. Moreover, in view of Theorem 2.1.1, if we wish to obtain a minimax, admissible estimator using this method, then we should also have $\phi(z) \leqq \phi^{(\circ)*}(z)$, for all z.

2.2. *Estimating one of a set of normal means with prior information.* Assume $X \sim MVN_p(\mu, \Sigma)$ where $\Sigma > 0$ is known. Let $\mu = (\mu_1, \mu_2, \cdots, \mu_p) \in R^p$. An estimate of μ_1 based on X is required, where the loss function is given by $L(\hat{\mu}_1; \mu_1) = W(\hat{\mu}_1 - \mu_1)$.

The problem remains invariant under the translation group, $\mathscr{G}$, whose action is described as

$$(2.2.1) \qquad X \to X + \beta\,, \qquad \mu \to \mu + \beta\,, \qquad \hat{\mu}_1 \to \hat{\mu}_1 + \beta_1\,,$$

where $\beta = (\beta_1, \beta_2, \cdots \beta_p) \in R^p$. Any $\mathscr{G}$-equivariant estimator has the form $X_1 + c$, and a risk function, $E_{\mu=0} W(X_1 + c)$, which is independent of μ. The function $c \to E_{\mu=0} W(X_1 + c)$ is easily shown to be strictly bowl-shaped, and so there is a best $\mathscr{G}$-equivariant estimator, $X_1 + c^{(\circ)}$, where

$$E_{\mu=0} W'(X_1 + c^{(\circ)}) = 0\,.$$

Now suppose it is known that $\mu_1 \leqq \mu_i$, $i = 2, 3, \cdots, p$. In this case it is possible to find an estimator of uniformly smaller risk than that of $X_1 + c^{(\circ)}$.

To this end let $\mathscr{H}$ denote the subgroup of $\mathscr{G}$ obtained from (2.2.1) by requiring that $\beta = \alpha(1, 1, \cdots, 1)$, for some scalar α. Any $\mathscr{H}$-equivariant estimator has the form $X_1 + \phi(Z)$, where $Z = (X_2 - X_1, \cdots, X_p - X_1)$. The risk of any such estimator, $r(\phi; \zeta)$, is a function of $\zeta = (\mu_2 - \mu_1, \cdots, \mu_p - \mu_1)$ alone, so that since only $\mathscr{H}$-equivariant estimators will be considered, we may assume $\mu_1 = 0$.

It is easily shown that $X_1 \mid Z = z \sim N[\sigma^2(\zeta - z)\delta', \sigma^2]$ while $Z \sim MVN_p[\zeta, \Delta^{-1}]$, where, if $D = \Sigma^{-1} = (d_{ij})$, $d_{i\bullet} = d_{\bullet i} = \Sigma_j d_{ij}$, $d_{\bullet\bullet} = \Sigma_i d_{i\bullet}$, and $D_{22} = (d_{ij})_{i,j>1}$, then $\sigma^2 = d_{\bullet\bullet}^{-1}$, $\delta = (d_{\bullet 2}, \cdots, d_{\bullet p})$, and $\Delta = D_{22} - \sigma^2\delta'\delta$.

Assume $d_{\bullet i} \geqq 0$, for all $i > 1$.

Let $\mathscr{C} = \{C(r) : r \leqq 0, r = \infty\}$, where $C(r) = \{z : \sigma^2 z\delta' \leqq r\}$. Using MLRP's in the manner of Lemmas 2.1.2 and 2.1.3, it can be shown that $c \to E_\zeta[W(X_1 + c) \mid Z = z]$ and $c \to E_\zeta[W(X_1 + c) \mid Z \in C]$ are strictly bowl-shaped for each $\zeta \in [0, \infty)^{p-1}$, $z \in R^{p-1}$, and $C \in \mathscr{C}$. Denote their respective minimizing values by $\phi_\zeta(z)$ and $\Psi_\zeta(C)$. Proceeding as in Section 2.1, we can show that $\phi_\zeta(z) \leqq \phi_0(z) = c^* + \sigma^2 z\delta'$, for all ζ, z, and for a constant c^* satisfying $EW'(X_1^* + c^*) = 0$, where $X_1^* \sim N(0, \sigma^2)$. Moreover, $\Psi_\zeta(C) \leqq \Psi_0(C)$ for all ζ, C and if C is a proper subset of D, $\Psi_0(C) < \Psi_0(D) \leqq c^{(\circ)}$.

For any estimator $X_1 + \phi(Z)$, let

(2.2.2) $$\phi^*(z) = \min\{\phi(z), \phi_0(z)\}.$$

Also, let

(2.2.3) $$\begin{aligned}\phi^{**}(z) &= c^{(\circ)}, & z\delta' > 0 \\ &= \Psi_0[C(\sigma^2 z\delta')], & z\delta' \leqq 0.\end{aligned}$$

We now present the counterparts of the main theorems of Section 1.

THEOREM 2.2.1. *For any* ζ

$$r(\phi^*; \zeta) \leqq r(\phi; \zeta)$$

with inequality if $P_\zeta[\phi^*(Z) \neq \phi(Z)] > 0$.

COROLLARY 2.2.1. *Let* $\phi^{(\circ)}(z) \equiv c^{(\circ)}$. *The risk function of* $X_1 + (\phi^{(\circ)})^*(Z)$ *is everywhere smaller than that of* $X_1 + c^{(\circ)}$.

THEOREM 2.2.2. *For any* ζ

$$r(\phi^{**}; \zeta) \leqq r(\phi^{(\circ)}; \zeta).$$

Moreover, $X_1 + \phi^{**}(Z)$ *is equal, for* $Z\delta' \leqq 0$, *to a generalized Bayes procedure within the class of* $\mathscr{H}$*-equivariant estimators; the prior is Lebesgue measure on the range of* ζ, $[0, \infty)^{p-1}$.

EXAMPLE 2.2.1. Suppose $W(u) = u^2$. Then $c^{(\circ)} = c^* = 0$, the best $\mathscr{G}$-equivariant estimator is X_1, and if $\mu_1 \leqq \mu_i$, $i = 1, 2, \cdots, p$ is known, a superior estimator is given by X_1 if $Z\delta' \geqq 0$ or $X_1 + \sigma^2 Z\delta'$ if $Z\delta' < 0$. The function ϕ^{**} defined in (2.2.3) is in this case

$$\begin{aligned}\phi^{**}(z) &= 0, & z\delta' \geqq 0 \\ &= -\sigma'/M(\sigma^2 z\delta'/\sigma'), & z\delta' < 0\end{aligned}$$

where $\sigma' = \sigma^2(\delta\Delta^{-1}\delta')^{\frac{1}{2}}$, $M(z) = \int_{-\infty}^{z} g(t)\,dt/g(z)$ and $g(t) = \exp(-\frac{1}{2}t^2)$.

REMARKS. The first of the superior alternatives to X_1, presented in Example 2.2.1 for quadratic loss, is, in a special case, a result of Srivastava and Bancroft (1967) which is discussed by Arnold (1970). The special case is obtained by requiring that $p = 2$ and that Σ be diagonal and the proofs given by these authors are different from those described here. The result in this special case can be described as follows: test $\mu_2 - \mu_1 = 0$ against $\mu_2 > \mu_1$ with $X_2 - X_1$ as the test statistic—use X_1 if the null hypothesis is rejected, and the pooled estimate, $(\sigma_1^{-2}X_1 + \sigma_2^{-2}X_2)(\sigma_1^{-2} + \sigma_2^{-2})^{-1}$, otherwise.

It is interesting that our analysis requires that $d_{i\cdot} \geqq 0$, $i > 1$. Actually this can be relaxed and instead the $d_{i\cdot}$ required to be of the same sign. Otherwise neither of our methods produces a superior alternative to X_1. The reason for this failure is most apparent when reconsidering the first method involving $\phi_\zeta(z)$, which is easily shown to be $c^* + \sigma^2(z - \zeta)\delta'$. Then, unless the $d_{i\cdot}$, $i > 1$, are of the same sign, $\zeta \to \phi_\zeta(z)$ is neither bounded above nor below and it is not possible for either $\{X_1 + c^{(\circ)} > X_1 + \phi_\zeta(Z) \text{ all } \zeta\}$ or $\{X_1 + c^{(\circ)} < X_1 + \phi_\zeta(Z) \text{ all } \zeta\}$ to occur with positive probability. So the method must fail. The condition that the $d_{i\bullet}$, $i > 1$, be of the same sign is not easy to interpret. When $p = 2$, it must hold. If $p = 3$, this condition can be stated in terms of the correlations, ρ_{ij}, between X_i and X_j. For example, $d_{2\bullet} < 0$ and $d_{3\bullet} > 0$ if and only if $\rho_{13} < \rho_{23}$, $\rho_{13} < \rho_{12}$ and $1 - \rho_{12} < \rho_{23} - \rho_{13}$. Thus if the X_i are quite dissimilarly correlated, for example, if ρ_{12} is near 0 while ρ_{13} is near -1 and ρ_{23} is near $+1$, the condition fails and the prior information $\mu_1 \leqq \mu_i$, $i = 2, 3$ cannot be used by our methods to produce superior alternatives to X_1.

Other forms of prior knowledge about ζ could be used in conjunction with our methods to produce superior alternatives to X_1. In the case of method 1, for example, if it is known that $\zeta \in \Lambda \subset (-\infty, \infty)^{p-1}$ and that either $\{\sup_{\zeta \in \Lambda} Z\delta' < \infty\}$ or $\{\inf_{\zeta \in \Lambda} Z\delta' > -\infty\}$ or both occur with positive probability, then improvement on X_1 is possible. We will not pursue this matter any further here.

It is possible to treat the problem of estimating one of several normal law variances, σ_i^2, $i = 1, \cdots, p$, in a manner analogous to our treatment in this section of the problem involving normal law means. It is not necessary to assume the means are known. We conclude our remarks by stating the answer obtained by applying the first of our methods to the problem of estimating σ_1^2 when the loss function is given by $L(\hat\sigma_1^2; \sigma_1^2, \cdots, \sigma_p^2, \mu_1, \cdots, \mu_p) = (\ln \hat\sigma_1^2\sigma_1^{-2})^2$.

Data consists of independent random variables $X_1, \cdots, X_p, T_1, \cdots T_p$ where $T_i\sigma_i^{-2}$ has the chi-squared distribution with n_i degrees of freedom and $X_i \sim N(\mu_i, \sigma_i^2)$. If χ_r^2 denotes the chi-squared random variable with r degrees of freedom, it is easily shown that the best fully equivariant estimator of σ_1^2 is given by $c^{(\circ)}T_1$, where $c^{(\circ)} = \exp[t(n_1)]$ and $t(r) = E \ln \chi_r^{-2}$. Assume now that $\sigma_1^2 \leqq \sigma_i^2$, $i > 1$ and consider the class of estimators having the form $\phi(Z, V)T_1$ where Z is as defined in Section 2.1 and $V = (1, T_2T_1^{-1}, \cdots, T_pT_1^{-1})$. For any estimator in this class let $\phi^*(Z, V) = \exp[\min\{\ln \phi(Z, V), t(n + p)\Sigma(V_i + Z_i^2)\}]$, where $n = n_1 + \cdots + n_p$. Then if $\sigma_1^2 \leqq \sigma_i^2$, $i > 1$, the risk function of $\phi^*(Z, V)T_1$ is never larger than that of $\phi(Z, V)T_1$. In particular, if $\phi^{(\circ)}(z, v) \equiv c^{(\circ)}$, it follows that $\phi^{(\circ)*}(Z, V)T_1$ has a uniformly smaller risk function than that of $c^{(\circ)}T_1$.

Added in proof. The choice of $\mathscr{C}$ in Theorem 2.2.2 was motivated by ϕ^*. If $\mathscr{C} = \{C(r): -\infty < r \leqq \infty\}$, then ϕ^{**} would be replaced by the generalized Bayes estimator $\Psi_0[C(\sigma^2 z\delta')]$. For admissibility considerations see Cohen and Sackrowitz [*Ann. Math. Statist.* **41** (1970) 2021–2034].

3. Techniques for improving estimators.

3.1. *Preliminaries.* $\mathscr{S}$ will denote in this section a possibly unbounded subset of R^m, $m \geqq 1$.

DEFINITION 3.1.1. $f: \mathscr{S} \to [0, \infty)$ is called symmetrical-bowl-shaped if there exists $m(f) \in \mathscr{S}$ and an inner product $(\cdot, \cdot)_f$ on R^m such that $(u - m(f), u - m(f)) > (v - m(f), v - m(f))$ implies $f(u) > f(v)$, $u, v \in \mathscr{S}$.

DEFINITION 3.1.2. If $\mathscr{S} \subset R$, $f: \mathscr{S} \to [0, \infty)$ is called strictly bowl-shaped if there exists $m(f) \in \mathscr{S}$ such that f is strictly decreasing (increasing) on $(-\infty, m(f)] \cap \mathscr{S}$ $(\mathscr{S} \cap [m(f), \infty))$.

The following lemma concerns any family, $\mathscr{F}$, of symmetrical-bowl-shaped functions on $\mathscr{S}$ for which the $(\cdot, \cdot)_f$ $f \in \mathscr{F}$ are identical; let $(\cdot, \cdot) = (\cdot, \cdot)_f$.

LEMMA 3.1. *Assume for a given* $w \in \mathscr{S}$ *there exists a hyperplane,* $\{x \in R^m: (x, a) = b\}$ *such that* $(w, a) > b$, *and* $(m(f), a) \leqq b$, $f \in \mathscr{F}$. *Then* $f(w) \geqq f(v_0)$, $f \in \mathscr{F}$ *where* $v_0 = w + (a, a)^{-1}[b - (w, a)]$, *if* $v_0 \in \mathscr{S}$.

PROOF. Observe that $(w - m(f), w - m(f)) > (v_0 - m(f), v_0 - m(f))$, $f \in \mathscr{F}$. The conclusion follows immediately from this observation.

3.2. *The statistical model.* $\mathscr{X}$, Θ, and $\mathscr{A}$ denote, respectively, the sample space, the parameter space, and the action space for the statistical problem under consideration. The observable random variable is X. A loss function $L: \mathscr{A} \times \Theta \to [0, \infty)$ is specified.

The problem remains invariant with respect to homomorphic transformation groups $\mathscr{G}$, $\bar{\mathscr{G}}$, and $\hat{\mathscr{G}}$ acting on $\mathscr{X}$, Θ, and $\mathscr{A}$ respectively. This means that $L(ga, \bar{g}\theta) \equiv L(a, \theta)$ and, if the distribution of X has parameter θ, that of gX has parameter $\bar{g}\theta$. Here $\bar{g} \in \bar{\mathscr{G}}$ and $\hat{g} \in \hat{\mathscr{G}}$ denote the homomorphic images of $g \in \mathscr{G}$.

To obtain a convenient representation of equivariant estimators, it is assumed that

$$\mathscr{X} = [\mathscr{G}] \times \mathscr{Z},$$

where $[\mathscr{G}] = \mathscr{G}/\mathscr{K}$ denotes the space of left cosets of some subgroup $\mathscr{K} \subset \mathscr{G}$ while $\mathscr{Z} = \mathscr{X}/\mathscr{G}$ denotes the space of $\mathscr{G}$-orbits in $\mathscr{X}$ (circumstances giving rise to such a representation are described, for example, by Koehn (1970)). $\mathscr{G}$ acts on $[\mathscr{G}]$; if $[g^*] \in [\mathscr{G}]$ labels $g^*\mathscr{K}$, $g[g^*]$ means $[gg^*]$. $\mathscr{G}$ leaves points of $\mathscr{Z}$ invariant.

The above representation of $\mathscr{X}$ enables us to write

$$X = (Y, Z), \qquad Y \in [\mathscr{G}], \qquad Z \in \mathscr{Z}.$$

$\bar{\mathscr{G}}$ is required to be transitive; for every pair $\theta_i \in \Theta$, $i = 1, 2$ there is at least one $\bar{g} \in \bar{\mathscr{G}}$ for which $\theta_1 = \bar{g}\theta_2$. As is easily shown, the marginal distribution of Z is known. Our analysis is conditional on Z; to simplify notation no explicit reference to Z is made.

An estimator $\phi: \mathscr{X} \to \mathscr{A}$ is called equivariant if $\phi(gx) = \hat{g}\phi(x)$ for all g and x. It follows that

$$\phi(x) = \phi([g]) = \hat{g}\phi([e])$$

where e is the identity transformation. But $\phi([e])$ is not arbitrary because $[g_1] = [g_2]$ implies $\hat{g}_1\phi([e]) = \hat{g}_2\phi([e])$, so that $\phi([e]) \in \mathscr{A}_0 = \{a: a \in A$ and $\mathscr{K}a = a\}$.

The risk function of ϕ,

$$r(\phi, \theta) = E_\theta L(\hat{G}\phi([e]), \theta)\,, \tag{3.2.1}$$

is easily shown to be independent of θ. It follows that $r(\phi, \theta)$ is a function of $\phi([e]) \in \mathscr{A}_0$ alone and if that function achieves its infimum, say at $a_0 \in \mathscr{A}_0$, an optimal equivariant estimator, $\phi_0(X) = \phi_0([G])$ is defined by

$$\phi_0([G]) = \hat{G}a_0\,. \tag{3.2.2}$$

ϕ_0 is well defined because $\hat{G}a_0$ depends on G only through $[G]$.

DEFINITION. The rule defined in equation (3.2.2) is called a best $\mathscr{G}$-equivariant estimator.

The search for a superior alternative to ϕ_0 is restricted to a subclass of estimators equivariant under a subgroup $\mathscr{H} \subset \mathscr{G}$. Let

$$[[\mathscr{G}]] = [\mathscr{G}]/\mathscr{H}$$

and

$$W = [[G]]$$

where $[[g]]$ labels the $\mathscr{H}$-orbit of $[g] \in [\mathscr{G}]$. The class of estimators considered are those of the form

$$\phi([G]) = \hat{G}\phi(W) \tag{3.2.3}$$

where $\phi: [[\mathscr{G}]] \to \mathscr{A}_0$.

It is easily shown that the risk function of any $\mathscr{H}$-equivariant estimator depends on θ only through $\omega = \omega(\theta)$, the maximal $\overline{\mathscr{H}}$-invariant labelling the $\overline{\mathscr{H}}$-orbits of Θ. Equivalently, the distribution of W depends only on ω.

3.3. *Main results.*

DEFINITION 3.3.1. Any subset, $\mathscr{C}$, of the σ-algebra of $[[\mathscr{G}]]$ is called a test-class. If $[[\mathscr{G}]] \in \mathscr{C}$ and $\mathscr{C}$ is both well-ordered and completely-ordered by inclusion, $\mathscr{C}$ is called an ordered test class.

For any test-class, $\mathscr{C}$, let $\mathscr{B}(\mathscr{C})$ denote the smallest σ-algebra containing $\mathscr{C}$. Another convenient notational device is $[x; y, z]$—which is defined as y, x, or z according as $x < y$, $y \leqq x \leqq z$, or $z < x$.

Define for all ω, w, $f_{\omega,w}(\cdot) = f_{\omega,w}(\cdot \mid \mathscr{C}): \mathscr{A}_0 \to [0, \infty)$ by

$$f_{\omega,w}(\phi) = E_\omega[L(\hat{G}\phi, \omega) \mid \mathscr{B}(\mathscr{C})](w)\,.$$

ASSUMPTION 1. $\mathscr{A}_0 \subset R$. For all ω, w, $f_{\omega,w}$ is strictly bowl-shaped on $\mathscr{A}_0$.

THEOREM 3.3.1. *Assume $\mathscr{C}$ is such that Assumption 1 holds. Then the risk function of $\hat{G}\phi^*(W) \equiv \hat{G}[a_0, \inf_\omega m(f_{\omega,W}), \sup_\omega (f_{\omega,W})]$ is never larger than that of ϕ_0 and smaller at any ω for which*

$$P_\omega[\phi^*(W) \neq a_0] > 0\,.$$

Suppose $\mathscr{A}_0 \subset R$ and $\mathscr{C}$ is an ordered test class. Let $C: [[\mathscr{G}]] \to \mathscr{C}$ be defined as

$$C(W) = \bigcap \{c : w \in c \in \mathscr{C}\}.$$

Assume $(\mathscr{C}, \mathscr{T})$ is a measurable space where $\mathscr{T}$ is a σ-algebra containing the intervals $(a, b] = \{c : a < c \leqq b, c \in \mathscr{C}\}$, $a, b \in \mathscr{C}$ chosen so that $C = C(W)$ is a random variable. Denote by $P_\omega{}^C$, the induced distribution of C.

We now consider as a class of possible alternatives to ϕ_0, estimators of the form

$$\phi([G]) = \hat{G}\psi(C), \qquad \psi(c) \in \mathscr{A}_0, \qquad c \in \mathscr{C}. \tag{3.3.1}$$

ψ is called approximable if there exists a double sequence $\{c(i, n); i = 1, \cdots, k_n, n = 1, 2, \cdots, c(k_n, n) \equiv [[\mathscr{G}]], c(i, n) \in \mathscr{C}\}$ such that

$$P_\omega{}^C[\psi(C) = \lim_{n\to\infty} \psi_n(C)] = 1, \qquad \text{all } \omega$$

where if I_A denotes the indicator function of A, $A \subset \mathscr{T}$,

$$\psi_n(c) = \sum_{i=1}^n \psi[c(i, n)] I_{(c(i-1,n),c(i,n)]}(c).$$

Define $g_{\omega,c}(\cdot \mid \mathscr{C}) : \mathscr{A}_0 \to [0, \infty)$, all $c \in \mathscr{C}$ and ω, by

$$g_{\omega,c}(\psi \mid \mathscr{C}) = E_\omega[L(\hat{G}\psi, \omega) \mid W \in c].$$

For simplicity, set $g_{\omega,c}(\cdot) = g_{\omega,c}(\cdot \mid \mathscr{C})$.

ASSUMPTION 2. $\mathscr{A}_0 \subset R$. $\mathscr{C}$ is an ordered test-class such that $P_\omega(W \in c) > 0$ for all $c \in \mathscr{C}$ and ω, $g_{\omega,c}(\cdot)$ is strictly bowl-shaped on $\mathscr{A}_0$ and either

(i) $\sup_\omega m(g_{\omega,\cdot})$ is approximable and strictly decreasing on $\mathscr{C}$ or
(ii) $\inf_\omega m(g_{\omega,\cdot})$ is approximable and strictly increasing on $\mathscr{C}$.

THEOREM 3.3.2. *Assume $\mathscr{C}$ is chosen so that Assumption 2 holds with condition* (i) [(ii)]. *Then the risk function of* $\hat{G} \sup_\omega m(g_{\omega,C})[\hat{G} \inf_\omega m(g_{\omega,C})]$ *is never larger than that of* ϕ_0.

ASSUMPTION 3. $\mathscr{A}_0 \subset R^m$. $f_{\omega,w}$ is symmetrical-bowl-shaped. $(\cdot, \cdot)_{F_{\omega,w}} = (\cdot, \cdot)$ all ω, w. For some $B \subset [[\mathscr{G}]]$, $w \in B$ implies that there exists a hyperplane, $\{x \in R^m : (x, a(w)) = b(w)$ such that $(a_0, a(w)) > b(w)$ and $(m(F_{\omega,w}), a(w)) \leqq b(w)$ all $\omega\}$.

If $\mathscr{C}$ is such that Assumption 3 holds, define ϕ^* by

$$\begin{aligned} \phi^*(w) &= a_0, & w \notin B, \\ &= a_0 + (a(w), a(w))^{-1}[b(w) - (a_0, a(w))]a(w), & w \in B. \end{aligned}$$

THEOREM 3.3. *Assume $\mathscr{C}$ is chosen so that Assumption 3 holds and that $\phi^*(w) \in \mathscr{A}_0$, for all w. Then the risk function of $\hat{G}\phi^*(W)$ is never larger than that of ϕ_0 and smaller at any ω such that $P_\omega(W \in B) > 0$.*

REMARKS. To achieve uniformity in our exposition certain simple extensions of our results were not included. For example, if prior knowledge restricts ω to a subset of its original range, the inadmissibility of the best equivariant estimator is anticipated. Obvious extensions of Theorems 3.1—3.3 might prove useful, as they did in Section 2.2, in finding an alternative.

Both Theorem 3.1 and Theorem 3.3 can be extended to yield conclusions about $\mathscr{H}$-equivariant estimators and Theorem 3.1 is used in its extended form in Section 2. Such an extension of Theorem 3.2 is not possible. $\phi_0([G]) = \hat{G}a_0$, rather than $\hat{G}\psi(W)$, is required, because then a_0, being constant, is measurable with respect to all sub-σ-algebras of $[[\mathscr{G}]]$.

For $m > 1$, it has been necessary to require symmetrical-bowl-shaped functions because of the geometric nature of the methods. When such symmetry is not present, it may be possible to look at a reduced one-dimensional problem, instead. For example, such a procedure was adopted by Brewster (1972) in interval estimation problems.

4. Acknowledgment. We are indebted to Dr. R. Shorrock for his helpful suggestions concerning the work of Section 2.1.

REFERENCES

[1] ARNOLD, B. C. (1970). An alternative derivation of a result due to Srivastava and Bancroft. *J. Roy. Statist. Soc. Ser. B* **32** 265–267.

[2] BREWSTER, J. F. (1972). On the admissibility of scale and quantile estimators. Ph. D. thesis, Univ. of British Columbia (unpublished).

[3] BROWN, L. D. (1968). Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters. *Ann. Math. Statist.* **39** 29–48.

[4] BROWN, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Statist.* **42** 855–903.

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

[6] JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 361–379. Univ. of California Press.

[7] KOEHN, U. (1970). Global cross sections and the densities of maximal invariants. *Ann. Math. Statist.* **41** 2045–2056.

[8] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

[9] SCLOVE, S. L., MORRIS, C. and RADHAKRISHNAN, R. (1972). Non-optimality of preliminary-test estimators for the mean of a multivariate normal distribution. *Ann. Math. Statist.* **43** 1481–1490.

[10] SRIVASTAVA, S. R. and BANCROFT, T. A. (1967). Inferences concerning a population correlation coefficient from one or possibly two samples subsequent to a preliminary test of significance. *J. Roy. Statist. Soc. Ser. B* **29** 282–291.

[11] STEIN, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.* **16** 155–160.

[12] STRAWDERMAN, W. E. (1972). Minimax estimation of powers of the variance of a normal population under squared error loss. Unpublished.

[13] ZIDEK, J. V. (1969). A representation of Bayes invariant procedures in terms of Haar measure. *Ann. Inst. Statist. Math.* **21** 291–308.

[14] ZIDEK, J. V. (1971). Inadmissibility of a class of estimators of a normal quantile. *Ann. Math. Statist.* **42** 1444–1447.

[15] ZIDEK, J. V. (1973). Estimating the scale parameter of the exponential distribution with unknown location. *Ann. Statist.* **1** 264–278.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MANITOBA
WINNIPEG, CANADA

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, CANADA