INADMISSIBILITY OF THE USUAL ESTIMATOR FOR THE MEAN OF A MULTIVARIATE NORMAL DISTRIBUTION

CHARLES STEIN
STANFORD UNIVERSITY

1. Introduction

If one observes the real random variables $X_1, \cdots, X_n$ independently normally distributed with unknown means $\xi_1, \cdots, \xi_n$ and variance 1, it is customary to estimate ξ_i by X_i. If the loss is the sum of squares of the errors, this estimator is admissible for $n \leqq 2$, but inadmissible for $n \geqq 3$. Since the usual estimator is best among those which transform correctly under translation, any admissible estimator for $n \geqq 3$ involves an arbitrary choice. While the results of this paper are not in a form suitable for immediate practical application, the possible improvement over the usual estimator seems to be large enough to be of practical importance if n is large.

Let X be a random n-vector whose expected value is the completely unknown vector ξ and whose components are independently normally distributed with variance 1. We consider the problem of estimating ξ with the loss function L given by

$$L(\xi, d) = (\xi - d)^2 = \Sigma(\xi_i - d_i)^2 \tag{1}$$

where d is the vector of estimates. In section 2 we give a short proof of the inadmissibility of the usual estimator

$$d = \xi_o(X) = X , \tag{2}$$

for $n \geqq 3$. For $n = 2$, the admissibility of ξ_o is proved in section 4. For $n = 1$ the admissibility of ξ_o is well known (see, for example, [1], [2], [3]) and also follows from the result for $n = 2$. Of course, all of the results concerning this problem apply with obvious modifications if the assumption that the components of X are independently distributed with variance 1 is replaced by the condition that the covariance matrix Σ of X is known and nonsingular and the loss function (1) is replaced by

$$L(\xi, d) = (\xi - d)'\Sigma^{-1}(\xi - d). \tag{3}$$

We shall give immediately below a heuristic argument indicating that the usual estimator ξ_o may be poor if n is large. With some additional precision, this could be made to yield a discussion of the infinite dimensional case or a proof that for sufficiently large n the usual estimator is inadmissible. We choose an arbitrary point in the sample space independent of the outcome of the experiment and call it the origin. Of course, in the way we have expressed the problem this choice has already been made, but in a correct coordinate-free presentation, it would appear as an arbitrary choice of one point in an affine space. Now

$$X^2 = (X - \xi)^2 + \xi^2 + 2\sqrt{\xi^2}Z \tag{4}$$

where

$$Z=\frac{\xi(X-\xi)}{\sqrt{\xi^2}} \tag{5}$$

has a univariate normal distribution with mean 0 and variance 1, and for large n, we have $(X-\xi)^2=n+O_p(\sqrt{n})$, so that

$$X^2=n+\xi^2+O_p(\sqrt{n+\xi^2}) \tag{6}$$

uniformly in ξ^2. (For the stochastic order notation o_p, O_p, see [4].) Consequently when we observe X^2 we know that ξ^2 is nearly X^2-n. The usual estimator ξ_0 would have us estimate ξ to lie outside of the convex set $\{\xi; \xi^2 \leqq X^2-cn\}$ (with c slightly less than 1) although we are practically sure that ξ lies in that set. It certainly seems more reasonable to cut X down at least by a factor of $[(X^2-n)/X^2]^{1/2}$ to bring the estimate within that sphere. Actually, because of the curvature of the sphere combined with the uncertainty of our knowledge of ξ, the best factor, to within the approximation considered here, turns out to be $(X^2-n)/X^2$. For, consider the class of estimators

$$\hat{\xi}(X)=\left[1-h\left(\frac{X^2}{n}\right)\right]X \tag{7}$$

where h is a continuous real-valued function with $\overline{\lim}_{t\to\infty}|t h(t)|<\infty$. We have (with $\rho^2=\xi^2/n$)

$$\begin{aligned}
[\hat{\xi}(X)-\xi]^2 &= \left\{\left[1-h\left(\frac{X^2}{n}\right)\right]X-\xi\right\}^2 \\
&= \left[1-h\left(\frac{X^2}{n}\right)\right]^2(X-\xi)^2+\xi^2h^2\left(\frac{X^2}{n}\right)-2\sqrt{\xi^2}h\left(\frac{X^2}{n}\right)\left[1-h\left(\frac{X^2}{n}\right)\right]Z \\
&= n\left[1-2h(1+\rho^2)+(1+\rho^2)h^2(1+\rho^2)\right]+O_p(\sqrt{n}).
\end{aligned} \tag{8}$$

This (without the remainder) attains its minimum of $n\rho^2/(1+\rho^2)$ for $h(1+\rho^2)=1/(1+\rho^2)$. In these calculations we have not used the normality.

In section 3 we consider some of the properties of spherically symmetric estimators, that is, estimators of the form (7), for finite n. We show that a spherically symmetric estimator is admissible provided it is admissible as compared with other spherically symmetric estimators. This is essentially a special case of a result given by Karlin [11] and Kudō [12].

In section 4 we use the information inequality in the manner of [1] and [2] in order to obtain lower bounds to the mean squared error of a spherical estimator of the mean. In particular, for $n=2$ this proves the admissibility of the usual estimator. For $n\geqq 3$ we obtain the bound $(n-2)^2/\xi^2$ for the asymptotic value of the possible improvement as $\xi^2\to\infty$, which is proved to be attainable in section 2.

In accordance with the results of section 3, a good spherically symmetric estimator is admissible for any n. However, roughly speaking as $n\to\infty$ it becomes less and less admissible, as in Robbins [7]. A simple way to obtain an estimator which is better for most practical purposes is to represent the parameter space (which is also essentially the sample space) as an orthogonal direct sum of two or more subspaces, also of large dimension and apply spherically symmetric estimators separately in each. If the ρ^2's (squared length of the population mean divided by the dimension) are appreciably different for the selected subspaces, this estimator will be better than the spherically symmetric one. It is unlikely that this estimator is admissible unless Bayes solutions (in the strict sense) are used in the component subspaces, but it is also unlikely that its departure from admissibility is important in practice.

In section 5 we consider very briefly a number of problems for which answers are needed before the methods of this paper can be applied with confidence.

2. Inadmissibility of the usual estimator

For $n \geq 3$, let X be a normally distributed random n-vector with unknown mean ξ and covariance matrix I, the identity matrix. In addition to the usual estimator ξ_o given by

$$\xi_o(X) = X \,, \tag{9}$$

we shall consider the estimator ξ_1 given by

$$\xi_1(X) = \left(1 - \frac{b}{a + X^2}\right) X \tag{10}$$

with $a, b > 0$. We shall show that for sufficiently small b and large a, ξ_1 is strictly better than ξ_o, in fact,

$$E_\xi [\xi_1(X) - \xi]^2 < n = E_\xi [\xi_o(X) - \xi]^2 \tag{11}$$

for all ξ. To prove (11) let $X = Y + \xi$ so that Y is normally distributed with mean 0 and covariance matrix I, the identity. Then

$$\begin{aligned} E_\xi\left[\left(1 - \frac{b}{a + X^2}\right) X - \xi\right]^2 &= E\left[Y - \frac{b}{a + (Y+\xi)^2}(Y+\xi)\right]^2 \\ &= n - 2bE\frac{Y(Y+\xi)}{a + (Y+\xi)^2} + b^2 E\frac{(Y+\xi)^2}{[a + (Y+\xi)^2]^2} \\ &< n - 2bE\frac{Y(Y+\xi) - b/2}{a + (Y+\xi)^2}. \end{aligned} \tag{12}$$

From the identity

$$\frac{1}{1+x} = 1 - x + \frac{x^2}{1+x} \tag{13}$$

we find that

$$\frac{1}{a + (Y+\xi)^2} = \frac{1}{a + Y^2 + \xi^2}\left\{1 - \frac{2\xi Y}{a + Y^2 + \xi^2} + \frac{4(\xi Y)^2}{(a + Y^2 + \xi^2)[a + (Y+\xi)^2]}\right\}. \tag{14}$$

Since the conditional mean of ξY given Y^2 is 0, we find from (14) that

$$E\frac{Y(Y+\xi) - b/2}{a + (Y+\xi)^2} = E\frac{Y^2 - b/2}{a + Y^2 + \xi^2} - 2E\frac{E[(\xi Y)^2 \mid Y^2]}{(a + Y^2 + \xi^2)^2} + 4E\frac{(\xi Y)^2[Y(Y+\xi) - b/2]}{(a + Y^2 + \xi^2)^2[a + (Y+\xi)^2]}. \tag{15}$$

But

$$E[(\xi Y)^2 \mid Y^2] = \frac{\xi^2}{n} Y^2 < \frac{a + Y^2 + \xi^2}{n} Y^2 \tag{16}$$

so that

$$(17)\quad E\frac{Y^2-b/2}{a+Y^2+\xi^2}-2E\frac{E[(\xi Y)^2|Y^2]}{(a+Y^2+\xi^2)^2}>E\frac{(1-2/n)Y^2-b/2}{a+Y^2+\xi^2}$$
$$>\frac{n-2}{n}E\frac{Y^2}{a+\xi^2}\left(1-\frac{Y^2}{a+\xi^2}\right)-\frac{b}{2(a+\xi^2)}$$
$$=\frac{n-2-b/2}{a+\xi^2}-\frac{(n-2)(n+2)}{(a+\xi^2)^2}.$$

It is intuitively clear that the last term on the right-hand side of (15) is $o[1/(a+\xi^2)]$ uniformly in ξ as $a\to\infty$. To give a detailed proof we observe that for $\xi^2\leqq a$

$$(18)\quad E\frac{(\xi Y)^3}{(a+Y^2+\xi^2)^2[a+(Y+\xi)^2]}\geqq -E\frac{|\xi Y|^3}{(a+Y^2+\xi^2)^2[a+(Y+\xi)^2]}$$
$$\geqq -E\frac{|\xi Y|^3}{(a+\xi^2)^2a}\geqq -\frac{c}{(a+\xi^2)a}.$$

For $\xi^2>a$,

$$(19)\quad E\frac{(\xi Y)^3}{(a+Y^2+\xi^2(^2[a+(Y+\xi)^2]}\geqq -E\frac{|\xi Y|^3}{(a+Y^2+\xi^2)^2[a+(Y+\xi)^2]}$$
$$\geqq -E\frac{|\xi Y|^3}{(a+Y^2+\xi^2)^2[a+\frac{1}{4}\xi^2]}$$
$$-P(\xi Y<-\tfrac{1}{4}\xi^2)E\left[\frac{|\xi Y|^3}{(a+\xi^2)^2a}\,\middle|\,\xi Y<-\tfrac{1}{4}\xi^2\right]$$
$$\geqq -\frac{c'}{(a+\xi^2)^{3/2}}-\frac{c''}{(a+\xi^2)a}.$$

Combining (18) and (19) we have

$$(20)\qquad E\frac{(\xi Y)^3}{(a+Y^2+\xi^2)^2[a+(Y+\xi)^2]}\geqq o\left(\frac{1}{a+\xi^2}\right)$$

uniformly in ξ as $a\to\infty$. Also

$$(21)\qquad E\frac{-\frac{b}{2}(\xi Y)^2}{(a+Y^2+\xi^2)^2[a+(Y+\xi)^2]}\geqq -\frac{b\xi^2}{2(a+\xi^2)^2a}=o\left(\frac{1}{a+\xi^2}\right)$$

uniformly in ξ as $a\to\infty$.

Thus from (12), (15), (17), (20), and (21) we find that

$$(22)\qquad E_\xi\left[\left(1-\frac{b}{a+X^2}\right)X-\xi\right]^2<n-2b\frac{n-2-b/2}{a+\xi^2}+o\left(\frac{1}{a+\xi^2}\right)$$

uniformly in ξ as $a\to\infty$. Consequently if we take $0<b<2(n-2)$, and a sufficiently large this will be less than n for all ξ. If we take $b=n-2$, then as $\xi^2\to\infty$, the improvement over the risk of the usual estimator is asymptotic to $(n-2)^2/\xi^2$. In section 4 we shall see that this is asymptotically the best possible improvement over the usual estimator in the neighborhood of $\xi^2=\infty$.

3. Spherically symmetrical estimators

We shall say that an estimator ξ is spherically symmetrical (about the origin) if it is of the form

$$\xi(x) = [1 - h(x^2)]\,x \tag{23}$$

where h is a real-valued function. This is equivalent to requiring that for every orthogonal transformation g, $g \circ \xi \circ g^{-1} = \xi$, that is, for all x

$$g[\xi(g^{-1}x)] = \xi(x). \tag{24}$$

First, if ξ is of the form (23), then

$$g[\xi(g^{-1}x)] = g\{[1 - h(x^2)]\,g^{-1}x\} = \xi(x). \tag{25}$$

Suppose conversely that ξ satisfies (24) for all orthogonal g. In particular, for those g which are reflections in a subspace containing x, $g[\xi(x)] = \xi(x)$. Consequently $\xi(x)$ lies along x, that is,

$$\xi(x) = [1 - h'(x)]\,x \tag{26}$$

for some real-valued function h'. Since a vector x can be taken into any other vector having the same squared length x^2 by an orthogonal transformation, this yields (23).

We shall show that if a spherically symmetric estimator ξ_2 is admissible as compared with all other spherically symmetric estimators, then it is admissible (in the class of all estimators). The proof is based on the compactness of the orthogonal group $\mathcal{G}$, and the continuity of the problem. It is similar to a proof for finite groups (see p. 228, [5], and p. 198, [6]). We shall only sketch the proof since a general result for compact groups will appear elsewhere. Because of the convexity of the loss function (1) in the estimate d we can confine our attention to nonrandomized procedures (see p. 186, [8]).

Suppose the estimator ξ is strictly better than the spherically symmetric estimator ξ_2, that is,

$$R_{\xi}(\xi) = E_{\xi}[\xi(X) - \xi]^2 \leqq E_{\xi}[\xi_2(X) - \xi]^2 \tag{27}$$

for all ξ with strict inequality for some ξ. Because of the continuity of R_{ξ} and R_{ξ_2}, strict inequality will hold for ξ in some nonempty open set S. Since ξ_2 is spherically symmetric, (27) will remain true if ξ is replaced by $g \circ \xi \circ g^{-1}$ with g orthogonal; in fact,

$$\begin{aligned} R_{g\circ\xi\circ g^{-1}}(\xi) &= E_{\xi}\{g[\xi(g^{-1}X)] - \xi\}^2 \\ &= E_{\xi}[\xi(g^{-1}X) - g^{-1}\xi]^2 = R_{\xi}(g^{-1}\xi). \end{aligned} \tag{28}$$

Thus, for fixed $\xi \in S$, the set of g for which $R_{g\circ\xi\circ g^{-1}}(\xi) < R_{\xi_2}(\xi)$ will be a nonempty open set. Let μ be the invariant probability measure on $\mathcal{G}$ which assigns strictly positive measure to any nonempty open set (for the existence of such a measure see chapter 2, [10]). Then

$$\xi' = \int g \circ \xi \circ g^{-1} d\mu(g) \tag{29}$$

is spherically symmetric, and because of the convexity of the loss function (1) in d

$$R_{\xi'}(\xi) \leqq \int R_{g\circ\xi\circ g^{-1}}(\xi)\, d\mu(g) \leqq R_{\xi_2}(\xi), \tag{30}$$

with strict inequality for $\xi \in S$. This shows that ξ_2 is not admissible in the class of all spherically symmetric estimators and completes the proof.

4. Application of the information inequality

In this section we apply the information inequality, as in [1] and [2], to obtain an upper bound for the possible improvement of a spherically symmetric estimator over the usual one. In particular, with the aid of the result of section 3, we show that for $n = 2$ the usual estimator is admissible.

Let $\hat{\xi}$ be any estimator of ξ with everywhere finite risk R and let b be the bias of $\hat{\xi}$, that is,

$$b(\xi) = E_\xi \hat{\xi}(X) - \xi . \tag{31}$$

Then by the information inequality

$$R(\xi) \geqq b^2(\xi) + \sum_i \Big\{ \sum_j \eta_{ij} [\delta_{ij} + b_{ij}(\xi)] \Big\}^2 \tag{32}$$

for any η with $\sum_j \eta_{ij}^2 = 1$ for all i, where $\delta_{ij} = 1$ if $i = j$, 0 otherwise and

$$b_{ij}(\xi) = \frac{\partial}{\partial \xi_j} b_i(\xi), \tag{33}$$

with $b_i(\xi)$ the ith coordinate of $b(\xi)$. Choosing

$$\eta_{ij} = \frac{\delta_{ij} + b_{ij}(\xi)}{\sqrt{\sum_j [\delta_{ij} + b_{ij}(\xi)]^2}} \tag{34}$$

so as to maximize the right-hand side of (32), we find

$$\begin{aligned} R(\xi) &\geqq b^2(\xi) + \sum_{i,j} [\delta_{ij} + b_{ij}(\xi)]^2 \\ &= b^2(\xi) + n + 2 \sum_i b_{ii}(\xi) + \sum_{i,j} b_{ij}^2(\xi) . \end{aligned} \tag{35}$$

In the spherically symmetrical case where $\hat{\xi}$ has the form (23), b has the form

$$b(\xi) = -\varphi(\xi^2)\xi \tag{36}$$

where φ is a differentiable real-valued function. In this case, dropping the last term, (35) becomes

$$R(\xi) \geqq n + \xi^2\varphi^2(\xi^2) - 2n\varphi(\xi^2) - 4\xi^2\varphi'(\xi^2) . \tag{37}$$

We first use (37) to prove that for $n = 2$ the usual estimator given by (2) is admissible. By the results of section 3, if $\hat{\xi}_o$ is not admissible there exists a spherically symmetric estimator $\hat{\xi}$ which is strictly better, and therefore there exists a function φ not vanishing identically such that

$$2 \geqq R(\xi) \geqq 2 + \xi^2\varphi^2(\xi^2) - 4\varphi(\xi^2) - 4\xi^2\varphi'(\xi^2) \tag{38}$$

for all $\xi^2 > 0$. Letting $t = \xi^2$ and $\psi(t) = t\varphi(t)$ we find

$$0 \geqq \frac{1}{t}\psi^2(t) - 4\psi'(t) \tag{39}$$

for $t > 0$. This shows that ψ is a nondecreasing function. We shall show that (39) implies that ψ is identically 0. Suppose first that $\psi(t_o) < 0$ for some $t_o > 0$. Then integrating the inequality

$$\frac{\psi'(t)}{\psi^2(t)} \geqq \frac{1}{4t} \tag{40}$$

from $t < t_o$ to t_o we obtain

$$-\frac{1}{\psi(t_o)} + \frac{1}{\psi(t)} \geqq \tfrac{1}{4} \log \frac{t_o}{t}. \tag{41}$$

The left-hand side is bounded as $t \to 0$ whereas the right-hand side approaches $+\infty$ so that this is a contradiction. If on the other hand $\psi(t_o) > 0$ for some $t_o > 0$, then

$$\frac{1}{\psi(t_o)} - \frac{1}{\psi(t)} \geqq \tfrac{1}{4} \log \frac{t}{t_o} \tag{42}$$

for all $t > t_o$. As $t \to \infty$, the left-hand side is bounded and the right-hand side approaches $+\infty$ so that we again have a contradiction.

Next we shall apply (35) to show that for $n \geqq 3$ there cannot exist $c > (n-2)^2$ and ξ_o^2 such that for all $\xi^2 \geqq \xi_o^2$,

$$R(\xi) \leqq n - \frac{c}{\xi^2}. \tag{43}$$

[We have seen in section 2 that there is an estimator which yields an improvement over the usual estimator asymptotic to $(n-2)^2/\xi^2$ as $\xi^2 \to \infty$.] It will suffice to show that the differential inequality

$$n - \frac{c}{\xi^2} \geqq n + \xi^2\varphi^2(\xi^2) - 2n\varphi(\xi^2) - 4\xi^2\varphi'(\xi^2), \tag{44}$$

obtained by combining (37) and (43) has no solution valid for all $\xi^2 \geqq \xi_o^2$. To see that (44) has no solution, let

$$\varphi(\xi^2) = \frac{n-2}{\xi^2} + f(\xi^2). \tag{45}$$

Then (44) becomes

$$-\frac{c-(n-2)^2}{\xi^2} \geqq \xi^2 f^2(\xi^2) - 4f(\xi^2) - 4\xi^2 f'(\xi^2). \tag{46}$$

Let $t = \xi^2$, $\psi(t) = tf(t)$. Then

$$-\frac{c-(n-2)^2}{t} \geqq \frac{1}{t}\psi^2(t) - 4\psi'(t), \tag{47}$$

$$\frac{\psi'(t)}{a^2+\psi^2(t)} \geqq \frac{1}{4t}, \tag{48}$$

where $a^2 = c - (n-2)^2$. From the inequality (39) (for all $t \geqq t_o$) which is weaker than (47) we concluded that $\psi(t) \leqq 0$. Consequently for $t_o < t$

$$\tan^{-1}\frac{\psi(t)}{a} - \tan^{-1}\frac{\psi(t_o)}{a} \geqq \frac{1}{4a}\log\frac{t}{t_o}. \tag{49}$$

The left-hand side is bounded (since ψ does not change sign) and the right-hand side approaches $+\infty$ as $t \to \infty$, which is a contradiction.

5. Miscellaneous remarks

In this section I shall indicate a few of the many problems which must be solved before the methods suggested in this paper can be applied with confidence in all situations where they seem appropriate.

(i) It seems that similar improvements must be possible if the variance is unknown, but there is available a reasonable number of degrees of freedom for estimating the variance. Presumably the correction to the sample mean will be smaller with a given estimated variance than if that value were known to be the variance. If there are no additional degrees of freedom for estimating the variance it is clear that the usual estimator is admissible. For, if there is a better estimator ξ, we can (because of the convexity of the loss function) construct a continuous estimator ξ' which is also better than ξ_o by taking, for example,

$$\xi'(x) = \frac{1}{(2\pi)^{n/2}} \int \xi(x-y)\, e^{-y^2/2} dy . \tag{50}$$

Then there is an $\epsilon > 0$ and a disc S of radius at least ϵ such that

$$[\xi'(x) - \xi_o(x)]^2 > \epsilon^2 \tag{51}$$

for all $x \in S$. If the variance of each component is much less than ϵ^2/n then the mean squared error of ξ_o will be small compared with ϵ^2 whereas that of ξ' will not.

(ii) The (positive definite) matrix of the quadratic loss function may be different from the inverse of the covariance matrix. It is intuitively clear that the usual estimator must be inadmissible provided there are at least three characteristic roots which do not differ excessively. However, because of the lack of spherical symmetry it seems difficult to select a good estimator.

(iii) The covariance matrix may be wholly or partially unknown. Suppose for example that the covariance matrix is completely unknown but there are enough degrees of freedom to estimate it. For simplicity suppose the matrix of the quadratic loss function is the inverse of the covariance matrix. Again it seems likely that the usual estimator is inadmissible. The problem of finding an admissible estimator better than the usual one is complicated by the fact that there is at present no reason to believe that the usual estimator of the covariance matrix is a good one.

(iv) At least two essentially different sequential problems suggest themselves. First we may consider observing, one at a time, random n-vectors $X^{(1)}, X^{(2)}, \cdots$ independently normally distributed with common unknown mean vector ξ and the identity matrix as covariance. If we want to attain a certain upper bound to the mean squared error with as small an expected number of observations as possible, it seems likely that we must resort to a sequential scheme.

Also, consider the situation where we observe, one at a time, real random variables $X_1, \cdots, X_n$ (with n fixed but large) independently normally distributed with unknown means $\xi_1, \cdots, \xi_n$ and variance 1. Suppose we want to estimate the ξ_i with the sum of squared errors as loss, but we are forced to estimate ξ_i immediately after observing X_i. It is not clear whether it is admissible to estimate ξ_i to be X_i. However, if it is admissible, it can only be because of the usual unrealistic assumption of a malevolent nature. In any case it should be possible to devise a scheme which will improve the estimate considerably if the means are small without doing much harm if they are large.

(v) It is not clear whether, in testing problems, the usual test may be inadmissible for the reasons given in this paper. This uncertainty cannot arise in the problem of testing for the variance of a normal distribution with unknown means as nuisance parameters (see p. 503, [9]). However, in the case of distinguishing between two possible values of the ratio of a mean to the standard deviation, with unknown means as nuisance parameters, the situation is unclear. Also the inadvisability (but not inadmissibility) of using spherical symmetry in a space of extremely high dimension is clear, at least if there is any natural way of breaking up the space.

(vi) It is of some interest to examine the situation in which the result of a previous experiment of the same type is taken as the origin. In this case (assuming the usual method, not that of this paper, has been applied to the previous experiment) ξ^2 is distributed as $\sigma^2\chi^2_n$, where σ^2 is the variance of each component in the first experiment so that for large n, $\xi^2 \sim \sigma^2 n$. Consequently the expected loss for the final estimate is nearly

$$\frac{\xi^2 n}{\xi^2+n} \approx \frac{n\sigma^2}{\sigma^2+1}, \tag{52}$$

which is the loss that would be attained if the two experiments were combined with weights inversely proportional to the variances. This method can be applied even if σ^2 is unknown.

(vii) Of course if we are interested in estimating only ξ_1 the presence of other unknown means $\xi_2, \cdots, \xi_n$ cannot make our task any easier. Thus our gain in over-all mean squared error must be accompanied by a deterioration in the mean squared error for certain components. Let us investigate this situation to the crudest approximation. We suppose without essential loss of generality that $\xi_1 = \sqrt{\xi^2}$, $\xi_2 = \cdots = \xi_n = 0$. Also we suppose n large and put $\rho^2 = \xi^2/n$. Then the best spherically symmetric estimator gives nearly the same result as $[1 - 1/(1 + \rho^2)]X$. Of course $X_1 = \sqrt{\xi^2} + Y_1 = \rho\sqrt{n} + Y_1$ where Y_1 is normally distributed with mean 0 and variance 1. The bias introduced in the estimate of the first coordinate is thus approximately $\rho\sqrt{n}/(1 + \rho^2)$ which makes a contribution of $n\rho^2/(1 + \rho^2)^2$ to the mean squared error of the estimate of this component. This attains its maximum of $n/4$ for $\rho = 1$. We notice that at this value of ρ, the squared errors of all other components combined add up to approximately the same amount $n/4$. For certain purposes this extreme concentration of the error in one component may be intolerable. This is one more reason why a space of extremely large dimension should be broken up before the methods of this paper are applied.

(viii) Better approximations than we have given here will be needed before this method can be applied to obtain simultaneous confidence sets for the means. Nevertheless it seems clear that we shall obtain confidence sets which are appreciably smaller geometrically than the usual discs centered at the sample mean vector.

(ix) For certain loss functions, for example

$$L(\xi, d) = \sup_i |\xi_i - d_i|, \tag{53}$$

little or no improvement over the usual estimator may be possible.

6. Acknowledgments

About two years ago I worked with L. J. Savage and E. Seiden on the problem of admissible strategies for the independent combination of two or more games. While no definite results were obtained, the discussion led to some clarification of the problem. I have also discussed the subject with J. Tukey and H. Robbins, who helped break down my conviction that the usual procedure must be admissible. Some remarks made by T. Harris and L. Moses when the paper was presented were also useful.

REFERENCES

[1] J. L. Hodges, Jr., and E. L. Lehmann, "Some applications of the Cramér-Rao inequality," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1951, pp. 13–22.

[2] M. A. Girshick and L. J. Savage, "Bayes and minimax estimates for quadratic loss functions," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1951, pp. 53–73.

[3] C. R. Blyth, "On minimax statistical decision procedures and their admissibility," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 22–42.

[4] H. B. Mann and A. Wald, "On stochastic limit and order relationships," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 217–226.

[5] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, New York, John Wiley and Sons, 1954, pp. 226–228.

[6] L. J. Savage, *The Foundations of Statistics*, New York, John Wiley and Sons, 1954.

[7] H. Robbins, "Asymptotically subminimax solutions of compound statistical decision problems," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1951, pp. 131–148.

[8] J. L. Hodges, Jr., and E. L. Lehmann, "Some problems in minimax point estimation," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 182–197.

[9] E. L. Lehmann and C. Stein, "Most powerful tests of composite hypotheses. I. Normal distributions," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 495–516.

[10] A. Weil, *L'Intégration dans les Groupes Topologiques et ses Applications*, Paris, Hermann, 1938.

[11] S. Karlin, "The theory of infinite games," *Annals of Math.*, Vol. 58 (1953), pp. 371–401.

[12] H. Kudō, "On minimax invariant estimates of the translation parameter," *Natural Science Report of the Ochanomizu University*, Vol. 6 (1955), pp. 31–73.