ESTIMATION WITH QUADRATIC LOSS

W. JAMES
FRESNO STATE COLLEGE
AND
CHARLES STEIN
STANFORD UNIVERSITY

1. Introduction

It has long been customary to measure the adequacy of an estimator by the smallness of its mean squared error. The least squares estimators were studied by Gauss and by other authors later in the nineteenth century. A proof that the best unbiased estimator of a linear function of the means of a set of observed random variables is the least squares estimator was given by Markov [12], a modified version of whose proof is given by David and Neyman [4]. A slightly more general theorem is given by Aitken [1]. Fisher [5] indicated that for large samples the maximum likelihood estimator approximately minimizes the mean squared error when compared with other reasonable estimators. This paper will be concerned with optimum properties or failure of optimum properties of the natural estimator in certain special problems with the risk usually measured by the mean squared error or, in the case of several parameters, by a quadratic function of the estimators. We shall first mention some recent papers on this subject and then give some results, mostly unpublished, in greater detail.

Pitman [13] in 1939 discussed the estimation of location and scale parameters and obtained the best estimator among those invariant under the affine transformations leaving the problem invariant. He considered various loss functions, in particular, mean squared error. Wald [18], also in 1939, in what may be considered the first paper on statistical decision theory, did the same for location parameters alone, and tried to show in his theorem 5 that the estimator obtained in this way is admissible, that is, that there is no estimator whose risk is no greater at any parameter point, and smaller at some point. However, his proof of theorem 5 is not convincing since he interchanges the order of integration in (30) without comment, and it is not clear that this integral is absolutely convergent. To our knowledge, no counterexample to this theorem is known, but in higher-dimensional cases, where the analogous argument seems at first glance only slightly less plausible, counterexamples are given in Blackwell [2] (which is discussed briefly at the end of section 3 of this paper) and in [14],

This work was supported in part by an ONR contract at Stanford University.

which is repeated in section 2 of this paper. In the paper of Blackwell an analogue of Wald's theorem 5 is proved for the special case of a distribution concentrated on a finite arithmetic progression. Hodges and Lehmann [7] proved some results concerning special exponential families, including Wald's theorem 5 for the special problem of estimating the mean of a normal distribution. Some more general results for the estimation of the mean of a normal distribution including sequential estimation with somewhat arbitrary loss function were obtained by Blyth [3] whose method is a principal tool of this paper. Girshick and Savage [6] proved the minimax property (which is weaker than admissibility here) of the natural estimator in Wald's problem and also generalized the results of Hodges and Lehmann to an arbitrary exponential family. Karlin [8] proved Wald's theorem 5 for mean squared error and for certain other loss functions under fairly weak conditions and also generalized the results of Girshick and Savage for exponential families. The author in [16] proved the result for mean squared error under weaker and simpler conditions than Karlin. This is given without complete proof as theorem 1 in section 3 of the present paper.

In section 2 of this paper is given a new proof by the authors of the result of [14] that the usual estimator of the mean of a multivariate normal distribution with the identity as covariance matrix is inadmissible when the loss is the sum of squares of the errors in the different coordinates if the dimension is at least three. An explicit formula is given for an estimator, still inadmissible, whose risk is never more than that of the usual estimator and considerably less near the origin. Other distributions and other loss functions are considered later in section 2. In section 3 the general problem of admissibility of estimators for problems with quadratic loss is formulated and a sufficient condition for admissibility is given and its relation to the necessary and sufficient condition [15] is briefly discussed. In section 4 theorems are given which show that under weak conditions Pitman's estimator for one or two location parameters is admissible when the loss is taken to be equal to the sum of squares of the errors. Conjectures are discussed for the more difficult problem where unknown location parameters are also present as nuisance parameters, and Blackwell's example is given. In section 5 a problem in multivariate analysis is given where the natural estimator is not even minimax although it has constant risk. These are related to the examples of one of the authors quoted by Kiefer [9] and Lehmann [11]. In section 6 some unsolved problems are mentioned.

The results of section 2 were obtained by the two authors working together. The remainder of the paper is the work of C. Stein.

2. Inadmissibility of the usual estimator for three or more location parameters

Let us first look at the spherically symmetric normal case where the inadmissibility of the usual estimator was first proved in [14]. Let X be a normally distributed p-dimensional coordinate vector with unknown mean $\xi = EX$ and covariance matrix equal to the identity matrix, that is, $E(X - \xi)(X - \xi)' = I$. We are interested in estimating ξ, say by $\hat{\xi}$ and define the loss to be

$$L(\xi, \hat{\xi}) = (\xi - \hat{\xi})'(\xi - \hat{\xi}) = ||\xi - \hat{\xi}||^2, \tag{1}$$

using the notation

$$||x||^2 = x'x. \tag{2}$$

The usual estimator is φ_0, defined by

$$\varphi_0(x) = x, \tag{3}$$

and its risk is

$$\rho(\xi, \varphi_0) = EL[\xi, \varphi_0(X)] = E(X - \xi)'(X - \xi) = p. \tag{4}$$

It is well known that among all unbiased estimators, or among all translation-invariant estimators (those φ for which $\varphi(x + c) = \varphi(x) + c$ for all vectors x and c), this estimator φ_0 has minimum risk for all ξ. However, we shall see that for $p \geqq 3$,

$$E\left|\left|\left(1 - \frac{p-2}{||X||^2}\right)X - \xi\right|\right|^2 = p - E\,\frac{(p-2)^2}{p - 2 + 2K} < p, \tag{5}$$

where K has a Poisson distribution with mean $||\xi||^2/2$. Thus the estimator φ_1 defined by

$$\varphi_1(X) = 1 - \frac{p-2}{||X||^2}\,X \tag{6}$$

has smaller risk than φ_0 for all ξ. In fact, the risk of φ_1 is 2 at $\xi = 0$ and increases gradually with $||\xi||^2$ to the value p as $||\xi||^2 \to \infty$. Although φ_1 is not admissible it seems unlikely that there are spherically symmetrical estimators which are appreciably better than φ_1. An analogous result is given in formulas (19) and (21) for the case where $E(X - \xi)(X - \xi)' = \sigma^2 I$, where σ^2 is unknown but we observe S distributed independently of X as σ^2 times a χ^2 with n degrees of freedom. For $p \leqq 2$ it is shown in [14] and also follows from the results of section 3 that the usual estimator is admissible.

We compute the risk of the estimator φ_2 defined by

$$\varphi_2(X) = \left(1 - \frac{b}{||X||^2}\right)X, \tag{7}$$

where b is a positive constant. We have

$$\begin{aligned} \rho(\xi, \varphi_2) &= E\left|\left|\left(1 - \frac{b}{||X||^2}\right)X - \xi\right|\right|^2 \\ &= E||X - \xi||^2 - 2bE\,\frac{(X-\xi)'X}{||X||^2} + b^2 E\,\frac{1}{||X||^2} \\ &= p - 2bE\,\frac{(X-\xi)'X}{||X||^2} + b^2 E\,\frac{1}{||X||^2}. \end{aligned} \tag{8}$$

It is well known that $||X||^2$, a noncentral χ^2 with p degrees of freedom and noncentrality parameter $||\xi||^2$, is distributed the same as a random variable W obtained by taking a random variable K having a Poisson distribution with mean $1/2||\xi||^2$ and then taking the conditional distribution of W given K to be that of a central χ^2 with $p + 2K$ degrees of freedom. Thus

$$E\frac{1}{||X||^2} = E\frac{1}{\chi^2_{p+2K}} = E\left(E\frac{1}{\chi^2_{p+2K}}\Big|K\right) = E\frac{1}{p-2+2K}. \tag{9}$$

To compute the expected value of the middle term on the right side of (8) let

$$U = \frac{\xi'X}{||\xi||}, \qquad V = \left\|X - \frac{\xi'X}{||\xi||^2}\xi\right\|^2. \tag{10}$$

Then

$$W = ||X||^2 = U^2 + V, \tag{11}$$

and U is normally distributed with mean $||\xi||$ and variance 1, and V is independent of U and has a χ^2 distribution with $p - 1$ d.f. The joint density of U and V is

$$\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(u - ||\xi||)^2\right\}\frac{1}{2^{(p-1)/2}\Gamma[(p-1)/2]}v^{(p-3)/2}e^{-v/2} \tag{12}$$

if $v \geqq 0$ and 0 if $v < 0$. Thus the joint density of U and W is

$$\frac{1}{\sqrt{2\pi}\,2^{(p-1)/2}\Gamma[(p-1)/2]}(w - u^2)^{(p-3)/2}\exp\left\{-\frac{1}{2}||\xi||^2 + ||\xi||\,u - \frac{1}{2}w\right\} \tag{13}$$

if $u^2 \leqq w$ and 0 elsewhere. It follows that

$$E\frac{U}{W} = \frac{\exp\left\{-\frac{1}{2}||\xi||^2\right\}}{\sqrt{2\pi}\,2^{(p-1)/2}\Gamma[(p-1)/2]}\int_0^\infty dw \int_{-\sqrt{w}}^{\sqrt{w}}\frac{u}{w}(w-u^2)^{(p-3)/2}\exp\left\{||\xi||u - \frac{1}{2}w\right\}. \tag{14}$$

Making the change of variable $t = u/\sqrt{w}$ we find

$$\begin{aligned}
&\int_0^\infty dw\int_{-\sqrt{w}}^{\sqrt{w}}\frac{u}{w}(w-u^2)^{(p-3)/2}\exp\left\{||\xi||u - \frac{1}{2}w\right\}du \\
&= \int_0^\infty w^{(p-3)/2}\exp\left\{-\frac{1}{2}w\right\}dw\int_{-1}^1 t(1-t^2)^{(p-3)/2}\exp\{||\xi||t\sqrt{w}\}\,dt \\
&= \int_0^\infty w^{(p-3)/2}\exp\left\{-\frac{1}{2}w\right\}dw\sum_{i=0}^\infty\frac{(||\xi||\sqrt{w})^i}{i!}\int_{-1}^1 t^{i+1}(1-t^2)^{(p-3)/2}\,dt \\
&= \sum_{j=0}^\infty\frac{||\xi||^{2j+1}}{(2j+1)!}\int_0^\infty w^{p/2+j-1}\exp\left\{-\frac{1}{2}w\right\}dw\int_{-1}^1 t^{2(j+1)}(1-t^2)^{(p-3)/2}\,dt
\end{aligned} \tag{15}$$

$$= \sum_{j=0}^{\infty} \frac{||\xi||^{2j+1}}{(2j+1)!} 2^{p/2+j}\Gamma\left(\frac{p}{2}+j\right) B\left(j+\frac{3}{2}, \frac{p-1}{2}\right)$$

$$= \sum_{j=0}^{\infty} \frac{||\xi||^{2j+1}\, 2^{p/2+j}\,\Gamma\left(j+\frac{3}{2}\right)\Gamma\left(\frac{p-1}{2}\right)}{(2j+1)!\left(\frac{p}{2}+j\right)}$$

$$= \Gamma\left(\frac{p-1}{2}\right) 2^{p/2} \sum_{j=0}^{\infty} \frac{2^j\Gamma\left(\frac{1}{2}\right)||\xi||^{2j+1}}{2^{2j+1}\Gamma(j+1)\left(\frac{p}{2}+j\right)}.$$

It follows from (10), (11), (14), and (15) that

$$(16)\qquad E\,\frac{(X-\xi)'X}{||X||^2} = 1 - ||\xi||E\,\frac{U}{W}$$

$$= 1 - ||\xi||\,\frac{\exp\left\{-\frac{1}{2}||\xi||^2\right\}}{\sqrt{\pi}} \sum_{j=0}^{\infty} \frac{\Gamma\left(\frac{1}{2}\right)||\xi||^{2j+1}}{2^{j+1}\Gamma(j+1)\left(\frac{p}{2}+j\right)}$$

$$= \exp\left\{-\frac{1}{2}||\xi||^2\right\}\left\{\sum_{i=0}^{\infty}\frac{\left(\frac{1}{2}||\xi||^2\right)^i}{i!} - ||\xi||^2\sum_{j=0}^{\infty}\frac{\left(\frac{1}{2}||\xi||^2\right)^j}{j!(p+2j)}\right\}$$

$$= \exp\left\{-\frac{1}{2}||\xi||^2\right\}\sum_{i=0}^{\infty}\frac{\left(\frac{1}{2}||\xi||^2\right)^i}{i!}\,\frac{p-2}{p-2+2i}$$

$$= (p-2)E\,\frac{1}{p-2+2K},$$

where K again has a Poisson distribution with mean $||\xi||^2/2$. Combining (8), (9), and (16) we find

$$(17)\qquad \rho(\xi,\varphi_2) = E\left\|\left(1-\frac{b}{||X||^2}\right)X - \xi\right\|^2$$

$$= p - 2(p-2)bE\,\frac{1}{p-2+2K} + b^2E\,\frac{1}{p-2+2K}.$$

This is minimized, for all ξ, by taking $b = p - 2$ and leads to the use of the estimator φ_1 defined by (6) and the formula (5) for its risk.

Now let us look at the case where X has mean ξ and covariance matrix given by

$$(18)\qquad E(X-\xi)(X-\xi)' = \sigma^2 I$$

and we observe S independent of X distributed as σ^2 times a χ^2 with n degrees of freedom. Both ξ and σ^2 are unknown. We consider the estimator φ_3 defined by

$$\varphi_3(X, S) = \left(1 - \frac{aS}{||X||^2}\right) X, \tag{19}$$

where a is a nonnegative constant. We have

$$\begin{aligned}\rho(\xi, \varphi_3) &= E||\varphi_3(X, S) - \xi||^2 \\ &= E||X - \xi||^2 - 2aE\frac{S(X - \xi)'X}{||X||^2} + a^2E\frac{S^2}{||X||^2} \\ &= \sigma^2\left\{p - 2aE\frac{S}{\sigma^2}E\frac{\left(\frac{X}{\sigma} - \frac{\xi}{\sigma}\right)'\frac{X}{\sigma}}{\left\|\frac{X}{\sigma}\right\|^2} + a^2E\left(\frac{S}{\sigma^2}\right)^2E\frac{1}{\left\|\frac{X}{\sigma}\right\|^2}\right\} \\ &= \sigma^2\left\{p - 2an(p - 2)E\frac{1}{p - 2 + 2K} + a^2n(n + 2)E\frac{1}{p - 2 + 2K}\right\}\end{aligned} \tag{20}$$

by (9) and (16) with X and ξ replaced by X/σ and ξ/σ respectively. Here K has a Poisson distribution with mean $||\xi||^2/2\sigma^2$. The choice $a = (p - 2)/(n + 2)$ minimizes (20) for all ξ, giving it the value

$$\rho(\xi, \varphi_3) = \sigma^2\left\{p - \frac{n}{n + 2}(p - 2)^2E\frac{1}{p - 2 + 2K}\right\}. \tag{21}$$

We can also treat the case where the covariance matrix is unknown but an estimate based on a Wishart matrix is available. Let X and S be independently distributed, X having a p-dimensional normal distribution with mean ξ and covariance matrix Σ and S being distributed as a $p \times p$ Wishart matrix with n degrees of freedom and expectation $n\Sigma$, where both ξ and Σ are unknown and Σ is nonsingular. Suppose we want to estimate ξ by $\hat{\xi}$ with loss function

$$L[(\xi, \Sigma), \hat{\xi}] = (\xi - \hat{\xi})'\Sigma^{-1}(\xi - \hat{\xi}). \tag{22}$$

We consider estimators of the form

$$\varphi(X, S) = \left(1 - \frac{c}{X'S^{-1}X}\right) X. \tag{23}$$

The risk function of φ is given by

$$\begin{aligned}\rho[(\xi, \Sigma), \varphi] &= E_{\xi,\Sigma}[\varphi(X, S) - \xi]'\,\Sigma^{-1}[\varphi(X, S) - \xi] \\ &= E_{\xi,\Sigma}\left[\left(1 - \frac{c}{X'S^{-1}X}\right)X - \xi\right]'\Sigma^{-1}\left[\left(1 - \frac{c}{X'S^{-1}X}\right)X - \xi\right] \\ &= E_{\xi^*,I}\left[\left(1 - \frac{c}{X'S^{-1}X}\right)X - \xi^*\right]'\left[\left(1 - \frac{c}{X'S^{-1}X}\right)X - \xi^*\right],\end{aligned} \tag{24}$$

where $\xi^{*\prime} = [(\xi'\Sigma^{-1}\xi)^{1/2}, 0, \cdots, 0]$. But it is well known (see, for example Wijsman [19]) that the conditional distribution of $X'S^{-1}X$ given X is that of $X'X/S^*$, where S^* is distributed as χ^2_{n-p+1} independent of X. Comparing (24) and (20) we see that the optimum choice of c is $(p-2)/(n-p+3)$ and, for this choice, the risk function is given by

$$\rho[(\xi, \Sigma), \varphi] = p - \frac{n-p+1}{n-p+3}(p-2)^2 E \frac{1}{p-2+2K}, \tag{25}$$

where K has a Poisson distribution with mean $(1/2)\xi'\Sigma^{-1}\xi$.

The improvement achieved by these estimators over the usual estimator may be understood better if we break up the error into its component along X and its component orthogonal to X. For simplicity we consider the case where the covariance matrix is known to be the identity. If we consider any estimator which lies along X, the error orthogonal to X is $\xi - (\xi'X/||X||^2)X$ and its mean square is

$$\begin{aligned} E_\xi\left\|\xi - \frac{\xi'X}{||X||^2}X\right\|^2 &= ||\xi||^2 - E_\xi\frac{(\xi'X)^2}{||X||^2} \\ &= ||\xi||^2\left(1 - E_\xi\frac{1+2K}{p+2K}\right) \\ &= (p-1)||\xi||^2 E_\xi\frac{1}{p+2K} \\ &= (p-1)\left[1 - (p-2)E_\xi\frac{1}{p-2+2K}\right]. \end{aligned} \tag{26}$$

Thus the mean square of the component along X of the error of $[1 - (p-2)/||X||^2]X$ is

$$\begin{aligned} &\left[p - (p-2)^2E_\xi\frac{1}{p-2+2K}\right] - (p-1)\left[1 - (p-2)E_\xi\frac{1}{p-2+2K}\right] \\ &\qquad = 1 + (p-2)E_\xi\frac{1}{p-2+2K} \leqq 2. \end{aligned} \tag{27}$$

On seeing the results given above several people have expressed fear that they were closely tied up with the use of an unbounded loss function, which many people consider unreasonable. We give an example to show that, at least qualitatively, this is not so. Again, for simplicity we suppose X a p-variate random vector normally distributed with mean ξ and covariance matrix equal to the identity matrix. Suppose we are interested in estimating ξ by $\hat{\xi}$ with loss function

$$L(\xi, \hat{\xi}) = F(||\hat{\xi} - \xi||^2), \tag{28}$$

where F has a bounded derivative and is continuouusly differentiable and concave (that is, $F''(t) \leqq 0$ for all $t > 0$). We shall show that for sufficiently small b and large a (independent of ξ) the estimator φ defined by

$$\varphi(X) = \left(1 - \frac{b}{a + ||X||^2}\right)X \tag{29}$$

has smaller risk than the usual estimator X. We have (with $Y = X - \xi$)

$$\begin{aligned}(30)\qquad \rho(\xi,\varphi) &= E_\xi F\left[\left\|\left(1-\frac{b}{a+\|X\|^2}\right)X-\xi\right\|^2\right]\\ &= E_\xi F\left[\|X-\xi\|^2-2b\frac{X'(X-\xi)}{a+\|X\|^2}+b^2\frac{\|X\|^2}{(a+\|X\|^2)^2}\right]\\ &\leqq E_\xi F(\|X-\xi\|^2)-2bE_\xi\left[\frac{X'(X-\xi)-\frac{b}{2}}{a+\|X\|^2}\right]F'(\|X-\xi\|^2)\\ &= EF(\|Y\|^2)-2bE\left[\frac{(Y+\xi)'Y-\frac{b}{2}}{a+\|Y+\xi\|^2}\right]F'(\|Y\|^2).\end{aligned}$$

Just as in [14] one can obtain by a Taylor expansion

$$\begin{aligned}(31)\qquad & E\left[\frac{(Y+\xi)'Y-\frac{b}{2}}{a+\|Y+\xi\|^2}\right]F'(\|Y\|^2)\\ &= E\frac{(Y+\xi)'Y-\frac{b}{2}}{a+\|\xi\|^2+\|Y\|^2}\left(1-\frac{2\xi'Y}{a+\|\xi\|^2+\|Y\|^2}\right)F'(\|Y\|^2)+o\left[\left(\frac{1}{a+\|\xi\|^2}\right)\right]\\ &= EE\left\{\frac{(Y+\xi)'Y-\frac{b}{2}}{a+\|\xi\|^2+\|Y\|^2}\left(1-\frac{2\xi'Y}{a+\|\xi\|^2+\|Y\|^2}\right)\middle|\,\|Y\|^2\right\}\\ &\qquad\qquad F'(\|Y\|^2)+o\left[\left(\frac{1}{a+\|\xi\|^2}\right)\right]\\ &= E\left\{\frac{\|Y\|^2-\frac{b}{2}}{a+\|\xi\|^2+\|Y\|^2}-\frac{2\|\xi\|^2\|Y\|^2/p}{(a+\|\xi\|^2+\|Y\|^2)^2}\right\}F'(\|Y\|^2)+o\left(\frac{1}{a+\|\xi\|^2}\right)\\ &\geqq E\frac{\left(1-\frac{2}{p}\right)\|Y\|^2-\frac{b}{2}}{a+\|\xi\|^2+\|Y\|^2}F'(\|Y\|^2)+o\left(\frac{1}{a+\|\xi\|^2}\right).\end{aligned}$$

To see that this is everywhere positive if b is sufficiently small and a sufficiently large we look at

$$(32)\qquad AE\frac{\|Y\|^2F'(\|Y\|^2)}{A+\|Y\|^2}=f(A)$$

and

$$(33)\qquad AE\frac{F'(\|Y\|^2)}{A+\|Y\|^2}=g(A).$$

It is clear that f and g are continuous strictly positive valued functions on $[1, \infty)$ with finite nonzero limits at ∞. It follows that

$$c = \inf_{A \geqq 1} f(A) = 0, \tag{34}$$

$$d = \sup_{A \geqq 1} g(A) < \infty. \tag{35}$$

Then, if b is chosen so that

$$\left(1 - \frac{2}{p}\right)c - \frac{b}{2}d > 0 \tag{36}$$

it follows from (30) and (31) that, for sufficiently large a

$$\rho(\xi, \varphi) < E_\xi F(\|X - \xi\|^2) \tag{37}$$

for all ξ.

The inadmissibility of the usual estimator for three or more location parameters does not require the assumption of normality. We shall give the following result without proof. Let X be a p-dimensional random coordinate vector with mean ξ, and finite fourth absolute moments:

$$E_\xi(X_i - \xi_i)^4 \leqq C < \infty. \tag{38}$$

For simplicity of notation we assume the X_i are uncorrelated and write

$$\sigma_i^2 = E_\xi(X_i - \xi_i)^2. \tag{39}$$

Then for $p \geqq 3$ and

$$b < 2(p - 2) \min \sigma_i^2 \tag{40}$$

and sufficiently large a depending only on C

$$E_\xi \sum \left[\left(1 - \frac{b}{\sigma_i^2[a + \sum X_i^2/\sigma_i^2]}\right) X_i - \xi_i\right]^2 < E_\xi \sum (X_i - \xi_i)^2 = \sum \sigma_i^2. \tag{41}$$

It would be desirable to obtain explicit formulas for estimators one can seriously recommend in the last two cases considered above.

3. Formulation of the general problem of admissible estimation with quadratic loss

Let $\mathcal{Z}$ be a set (the sample space), $\mathcal{B}$ a σ-algebra of subsets of $\mathcal{Z}$ and λ a σ-finite measure on $\mathcal{B}$. Let Θ be another set (the parameter space), $\mathcal{C}$ a σ-algebra of subsets of Θ, and $p(\cdot|\cdot)$ a nonnegative valued $\mathcal{BC}$-measurable function on $\mathcal{Z} \times \Theta$ such that for each $\theta \in \Theta$, $p(\cdot|\theta)$ is a probability density with respect to λ, that is,

$$\int p(z|\theta)\, d\lambda(z) = 1. \tag{42}$$

Let A, the action space, be the k-dimensional real coordinate space, α a $\mathcal{C}$-measurable function on Θ to the set of positive semidefinite symmetric $k \times k$ matrices, and η a $\mathcal{C}$-measurable function on Θ to A. We observe Z distributed over $\mathcal{Z}$ according to the probability density $p(\cdot|\theta)$ with respect to λ, where θ is unknown, then choose an action $a \in A$ and suffer the loss

$$L(\theta, a) = [a - \eta(\theta)]'\alpha(\theta)[a - \eta(\theta)]. \tag{43}$$

An estimator φ of $\eta(\theta)$ is a $\mathcal{B}$-measurable function on $\mathcal{Z}$ to A, the interpretation being that after observing Z one takes action $\varphi(Z)$, or in other words, estimates $\eta(\theta)$ to be $\varphi(Z)$. The risk function $\rho(\cdot, \varphi)$ is the function on Θ defined by

$$\rho(\theta, \varphi) = E_\theta[\varphi(Z) - \eta(\theta)]'\alpha(\theta)[\varphi(Z) - \eta(\theta)]. \tag{44}$$

Roughly speaking, we want to choose φ so as to keep $\rho(\cdot, \varphi)$ small, but this is not a precise statement since, for any given φ it will usually be possible to modify φ so as to decrease $\rho(\theta, \varphi)$ at some θ but increase it at other θ. In many problems there is a commonly used estimator, for example, the maximum likelihood estimator or one suggested by invariance, linearity, unbiasedness, or some combination of these. Then it is natural to ask whether this estimator is admissible in the sense of Wald. The estimator φ_0 is said to be admissible if there is no estimator φ for which

$$\rho(\theta, \varphi) \leqq \rho(\theta, \varphi_0) \tag{45}$$

for all θ with strict inequality for some θ. If there does exist such a φ then φ is said to be better than φ_0 and φ_0 is said to be inadmissible. We shall also find it useful to define an estimator φ_0 to be almost admissible with respect to a measure Π on the σ-algebra $\mathcal{C}$ of subsets of Θ if there is no estimator φ for which (45) holds for all θ with strict inequality on a set having positive Π-measure.

Next we give a simple sufficient condition for almost-admissibility of certain estimators. Although we do not discuss the necessity of the condition, its similarity to the necessary and sufficient condition of [15] leads us to apply it with confidence. It should be remarked that in [15] the condition of boundedness of the risk of the estimator (or strategy, b_0 in the notation of that paper) was inadvertently omitted in theorems 3 and 4. It is needed in order to justify the reduction to (48). If Π is a σ-finite measure on $\mathcal{C}$ we shall define

$$\varphi_\Pi(x) = \left[\int \alpha(\theta)p(x|\theta)\, d\Pi(\theta)\right]^{-1} \int \alpha(\theta)\eta(\theta)p(x|\theta)\, d\Pi(\theta) \tag{46}$$

provided the integrals involved are finite almost everywhere. Observe that if Π is a probability measure, φ_Π is the Bayes' estimator of $\eta(\theta)$, that is, $\varphi = \varphi_\Pi$ minimizes

$$\int d\Pi(\theta)E_\theta[\varphi(X) - \eta(\theta)]'\alpha(\theta)[\varphi(X) - \eta(\theta)]. \tag{47}$$

If q is a probability density with respect to Π we shall write (q, Π) for the induced probability measure, that is,

$$(q, \Pi)S = \int_S q(\theta)\, d\Pi(\theta). \tag{48}$$

THEOREM 3.1. *If φ is an estimator of $\eta(\theta)$ with bounded risk such that for each set C in a denumerable family $\mathfrak{F}$ of sets whose union is Θ*

$$\inf_{q \in \mathcal{S}(C)} \frac{\int q(\theta)\, d\Pi(\theta) E_\theta[\varphi(X) - \varphi_{(q,\Pi)}(X)]'\alpha(\theta)[\varphi(X) - \varphi_{(q,\Pi)}(X)]}{\int_C q(\theta)\, d\Pi(\theta)} = 0, \tag{49}$$

where $\mathcal{S}(C)$ is the set of probability densities with respect to Π which are constant (but not 0) on C, then φ is almost admissible with respect to Π.

PROOF. Suppose to the contrary that φ_1 is such that

$$E_\theta[\varphi_1(X) - \eta(\theta)]'\alpha(\theta)[\varphi_1(X) - \eta(\theta)] \leqq E_\theta[\varphi(X) - \eta(\theta)]'\alpha(\theta)[\varphi(X) - \eta(\theta)] \tag{50}$$

with strict inequality on a set S of positive Π-measure. Choose $\epsilon > 0$ and C a set in $\mathfrak{F}$ such that $\Pi(C \cap S_\epsilon) > 0$ where S_ϵ is the set of θ for which

$$E_\theta[\varphi_1(X) - \eta(\theta)]'\alpha(\theta)[\varphi_1(X) - \eta(\theta)] \leqq E_\theta[\varphi(X) - \eta(\theta)]'\alpha(\theta)[\varphi(X) - \eta(\theta)] - \epsilon. \tag{51}$$

Then, for any $q \in \mathcal{S}(C)$

$$\int q(\theta)\, d\Pi(\theta) E_\theta[\varphi_1(X) - \eta(\theta)]'\alpha(\theta)[\varphi_1(X) - \eta(\theta)] \leqq \int q(\theta)\, d\Pi(\theta) E_\theta[\varphi(X) - \eta(\theta)]'\alpha(\theta)[\varphi(X) - \eta(\theta)] - \epsilon \int_{C \cap S_\epsilon} q(\theta)\, d\Pi(\theta). \tag{52}$$

It follows that

$$\begin{aligned} \frac{\epsilon \Pi(C \cap S_\epsilon)}{\Pi(C)} &= \epsilon \frac{\int_{C \cap S_\epsilon} q(\theta)\, d\Pi(\theta)}{\int_C q(\theta)\, d\Pi(\theta)} \\ &\leqq \frac{1}{\int_C q(\theta)\, d\Pi(\theta)} \Big\{ \int q(\theta)\, d\Pi(\theta) E_\theta[\varphi(X) - \eta(\theta)]'\alpha(\theta)[\varphi(X) - \eta(\theta)] \\ &\qquad - \int q(\theta)\, d\Pi(\theta) E_\theta[\varphi_1(X) - \eta(\theta)]'\alpha(\theta)[\varphi_1(X) - \eta(\theta)] \Big\} \\ &\leqq \frac{1}{\int_C q(\theta)\, d\Pi(\theta)} \Big\{ \int q(\theta)\, d\Pi(\theta) E_\theta[\varphi(X) - \eta(\theta)]'\alpha(\theta)[\varphi(X) - \eta(\theta)] \\ &\qquad - \int q(\theta)\, d\Pi(\theta) E_\theta[\varphi_{(q,\Pi)}(X) - \eta(\theta)]'\alpha(\theta)[\varphi_{(q,\Pi)}(X) - \eta(\theta)] \Big\} \\ &= \frac{\int q(\theta)\, d\Pi(\theta) E_\theta[\varphi(X) - \varphi_{(q,\Pi)}(X)]'\alpha(\theta)[\varphi(X) - \varphi_{(q,\Pi)}(X)]}{\int_C q(\theta)\, d\Pi(\theta)} \end{aligned} \tag{53}$$

and this contradicts (49).

4. Admissibility of Pitman's estimator for location parameters in certain low dimensional cases

The sample space $\mathcal{Z}$ of section 3 is now of the form $\mathfrak{X} \times \mathcal{Y}$, where $\mathfrak{X}$ is a finite-dimensional real coordinate space, and $\mathcal{Y}$ arbitrary and the σ-algebra $\mathcal{B}$ is a product σ-algebra $\mathcal{B} = \mathcal{B}_1\mathcal{B}_2$ where $\mathcal{B}_1$ consists of the Borel sets in $\mathfrak{X}$ and $\mathcal{B}_2$ is an arbitrary σ-algebra of subsets of $\mathcal{Y}$. Here λ is the product measure $\lambda = \mu_\nu$ where μ is a Lebesgue measure on $\mathcal{B}_1$ and ν is an arbitrary probability measure on $\mathcal{B}_2$. The parameter space Θ and the action space A coincide with $\mathfrak{X}$. The loss function is

$$L(\theta, a) = (a - \theta)'(a - \theta). \tag{54}$$

We observe (X, Y) whose distribution, for given θ, is such that Y is distributed according to ν and the conditional density of $X - \theta$ given Y is $p(\cdot|Y)$, a known density. We assume

$$\int p(x, y)\, dx = 1 \tag{55}$$

and

$$\int xp(x, y)\, dx = 0 \tag{56}$$

for all y. Condition (56) is introduced only for the purpose of making the natural estimator X (see [6] or [16]). The condition (49) for the natural estimator X to be almost admissible becomes the existence of a sequence π_σ of densities with respect to Lebesgue measure in $\mathfrak{X}$ such that

$$\lim_{\sigma\to\infty} \frac{1}{\inf\limits_{||x||\leqq 1} \pi_\sigma(x)} \int d\nu(y) \int dx \frac{\left\|\int \eta\pi_\sigma(x - \eta)p(\eta|y)\, d\eta\right\|^2}{\int \pi_\sigma(x - \eta)p(\eta|y)\, d\eta} = 0. \tag{57}$$

This is derived in formula (63) below.

In [16] one of the authors proved the following theorem.

THEOREM 4.1. *When* $\dim \mathfrak{X} = 1$, *if in addition to the above conditions*

$$\int d\nu(y) \left[\int x^2 p(x|y)\, dx\right]^{3/2} < \infty, \tag{58}$$

then X is an admissible estimator of θ, that is, there does not exist a function φ such that

$$\int d\nu(y) \int [\varphi(x, y) - \theta]^2 p(x - \theta|y)\, dx \leqq \int d\nu(y) \int x^2 p(x|y)\, dx \tag{59}$$

for all θ with strict inequality for some θ.

This is proved by first showing that (57) holds with

$$\pi_\sigma(x) = \frac{1}{\pi\sigma\left(1 + \frac{x^2}{\sigma^2}\right)}, \tag{60}$$

so that X is almost admissible, and then proving that this implies that X is admissible. It is not clear that the condition (58) is necessary.

We shall sketch the proof of a similar but more difficult theorem in the two-dimensional case.

THEOREM 2. *When* $\dim \mathfrak{X} = 2$, *if in addition to* (55) *and* (56)

$$\int d\nu(y) \left[\int ||x||^2 \log^{1+\delta} ||x||^2 p(x, y)\, dx\right]^2 < \infty, \tag{61}$$

then X is an admissible estimator of θ, that is, there does not exist a function φ on $\mathfrak{X} \times \mathfrak{Y}$ to $\mathfrak{X}$ such that

$$\int d\nu(y) \int ||\varphi(x, y) - \theta||^2 p(x - \theta|y)\, dx \leqq \int d\nu(y) \int ||x||^2 p(x|y)\, dx \tag{62}$$

for all θ with strict inequality for some θ.

SKETCH OF PROOF. First we show that (57) implies (49) in the present case. We take Π to be a Lebesgue measure in $\mathfrak{X}$ and in accordance with the notation of the present section write π_σ instead of q. Because of the invariance of our problem under translation we need prove (49) only for C equal to a solid sphere of radius 1 about the origin. Then the integral in the numerator of (49) is

$$\int \pi_\sigma(\xi)\, d\xi E_\xi \left\| X - \frac{\int \xi_1 p(X - \xi_1, Y)\pi_\sigma(\xi_1)\, d\xi_1}{\int p(X - \xi_1, Y)\pi_\sigma(\xi_1)\, d\xi_1} \right\|^2 \tag{63}$$

$$= \int \pi_\sigma(\xi)\, d\xi \int d\nu(y) \int p(x - \xi, y)\, dx \frac{\left\| \int (x - \xi_1) p(x - \xi_1, y)\pi_\sigma(\xi_1)\, d\xi_1 \right\|^2}{\left[\int p(x - \xi_1, y)\pi_\sigma(\xi_1)\, d\xi_1\right]^2}$$

$$= \int d\nu(y) \int dx \int \pi_\sigma(\xi)\, d\xi\, p(x - \xi, y) \frac{\left\| \int (x - \xi_1) p(x - \xi_1, y)\pi_\sigma(\xi_1)\, d\xi_1 \right\|^2}{\left[\int p(x - \xi_1, y)\pi_\sigma(\xi_1)\, d\xi_1\right]^2}$$

$$= \int d\nu(y) \int dx \frac{\left\| \int (x - \xi) p(x - \xi, y)\pi_\sigma(\xi)\, d\xi \right\|^2}{\int p(x - \xi, y)\pi_\sigma(\xi)\, d\xi}$$

$$= \int d\nu(y) \int dx \frac{\left\| \int \eta \pi_\sigma(x - \eta) p(\eta, y)\, d\eta \right\|^2}{\int \pi_\sigma(x - \eta) p(\eta, y)\, d\eta}$$

and it is clear that (57) implies (49).

To prove theorem 2 we define π_σ by

$$\pi_\sigma(\xi) = \begin{cases} \dfrac{K_\sigma}{\sigma} \log^2 \sqrt{\sigma}, & \|\xi\|^2 \leqq 1, \\ \dfrac{K_\sigma}{\sigma} \log^2 \left(\dfrac{\sqrt{\sigma}}{\|\xi\|}\right), & 1 \leqq \|\xi\|^2 \leqq M\sigma, \\ \dfrac{K_\sigma}{\sigma} \dfrac{B}{\dfrac{\|\xi\|^2}{\sigma} \log^\beta \left(\dfrac{A\|\xi\|}{\sqrt{\sigma}}\right)}, & \|\xi\|^2 \geqq M\sigma, \end{cases} \tag{64}$$

where $1 < \beta < 1 + \delta$, $0 < M < 1$, $A\sqrt{M} > 1$, and these constants and B are chosen so that π is continuous everywhere and continuously differentiable except at $\|\xi\|^2 = 1$. A computation analogous to that in [16] but much more tedious leads to the following bound for the inner integral on the right side of (63)

$$\begin{aligned} I(y) &= \int dx \frac{\left\|\int \eta\pi_\sigma(x - \eta)p(\eta, y)\, d\eta\right\|^2}{\int \pi_\sigma(x - \eta)p(\eta, y)\, d\eta} \\ &\leqq C_1 \frac{\log \sigma}{\sigma} \left[\int \eta^2(1 + \log^\beta \eta^2)p(\eta, y)\, d\eta\right]^2 \end{aligned} \tag{65}$$

if

$$\int \eta^2 p(\eta, y)\, d\eta \leqq C_2\sigma \tag{66}$$

where C_1 and C_2 are positive constants. Also by applying (63) with $\mathcal{Y}$ reduced to the point y it is easy to see that

$$I(y) \leqq \int \eta^2 p(\eta, y)\, d\eta \qquad \text{for all } y. \tag{67}$$

Now let S_σ be the set of y for which

$$\int \eta^2(1 + \log^\beta \eta^2)p(\eta, y)\, d\eta \leqq C_2\sigma. \tag{68}$$

Then

$$\begin{aligned} \int d\nu(y) \int dx & \frac{\left\|\int \eta\pi_\sigma(x - \eta)p(\eta, y)\, d\eta\right\|^2}{\int \pi_\sigma(x - \eta)p(\eta, y)\, d\eta} \\ &\leqq \int_{S_\sigma} d\nu(y) C_1 \frac{\log \sigma}{\sigma} \left[\int \eta^2(1 + \log^\beta \eta^2)p(\eta, y)\, d\eta\right]^2 \\ &\quad + \int_{\bar{S}_\sigma} d\nu(y) \int \eta^2(1 + \log^\beta \eta^2)p(\eta, y)\, d\eta. \end{aligned} \tag{69}$$

But

$$\begin{aligned} \int_{\bar{S}_\sigma} d\nu(y) & \int \eta^2(1 + \log^\beta \eta^2)p(\eta, y)\, d\eta \\ &\leqq \frac{1}{C_2\sigma} \int_{\bar{S}_\sigma} d\nu(y) \left[\int \eta^2(1 + \log^\beta \eta^2)p(\eta, y)\, d\eta\right]^2. \end{aligned} \tag{70}$$

It follows from (69) and (70) and (64) that

$$(71)\qquad \frac{1}{\pi_\sigma(0)}\int d\nu(y)\,dx\,\frac{\left\|\int \eta\pi_\sigma(x-\eta)p(\eta,y)\,d\eta\right\|^2}{\int \pi_\sigma(x-\eta)p(\eta,y)\,d\eta} \leqq C_3\frac{\log\sigma/\sigma}{\log^2\sigma/\sigma}\to 0 \quad \text{as} \quad \sigma\to\infty$$

so that theorem 2 follows from the remarks around (57).

Theorems 1 and 2 together with the results of section 2 settle in a fairly complete manner the question of admissibility of Pitman's estimator of a number of location parameters with positive definite translation-invariant quadratic loss function. If enough moments exist, Pitman's estimator is admissible if the number of parameters to be estimated is one or two, but inadmissible if this number is at least three. In the case where there are also parameters which enter as nuisance parameters, the only known results are an example of Blackwell [2] and some trivial consequences of the results given above.

Blackwell formulates his example as a somewhat pathological case involving one unknown location parameter where Pitman's estimator is almost admissible in the class of Borel-measurable estimators but not admissible (and not even almost admissible if measurability is not required). However, it can be reformulated as a nonpathological example involving four unknown location parameters with a quadratic loss function of rank one.

Now consider the problem where we have p unknown location parameters and the positive semidefinite translation-invariant quadratic loss function has rank r. From the result given without proof at the end of section 2, it follows that if $r \geqq 3$ and all fourth absolute moments exist, then Pitman's estimator is inadmissible. This follows from the application of the result at the end of section 2 to the problem we obtain if we look only at Pitman's estimator of the r parameters which enter effectively into the loss function, ignoring the rest of the observation. If $p = 2$ and $r = 1$, then, subject to the conditions of theorem 2, Pitman's estimator is admissible. If it were not, we could obtain a better estimator than Pitman's for a problem with $r = 2$ (contradicting theorem 2) by using the better estimator for the parameter which occurs in the original loss function of rank 1 and Pitman's estimator for another linearly independent parameter, with the new loss function equal to the sum of squares of the errors in estimating the parameters defined in this way. If r is 1 or 2 but p arbitrary, Pitman's estimator is admissible (subject to the existence of appropriate moments) if, for some choice of the coordinate system defining the $r - p$ nuisance parameters, Pitman's estimator coincides with what would be Pitman's estimator if these $r - p$ nuisance parameters were known. This is always the case if $r \leqq 2$ and the observed random vector X is normally distributed.

These are all the results known to me for the problem of estimating unknown location parameters with positive semidefinite translation-invariant quadratic loss function. In the following conjectures it is to be understood that in all cases sufficiently many moments are assumed to exist. I conjecture that if $p = 3$ and $r = 1$, in the case of a single orbit (that is, when $\mathcal{Y}$, analogous to that of theorems 1 and 2, reduces to a single point) Pitman's estimator is admissible, but this does not hold in general when $\mathcal{Y}$ does not reduce to a point. In the other cases not covered in the preceding paragraph, that is, if $p \geqq 3$ and $r = 2$ or if $p \geqq 4$ and $r = 1$, I conjecture that Pitman's estimator is, in general, inadmissible, but of course there are many exceptions, in particular those mentioned at the end of the last paragraph. Blackwell's example supports this conjecture for $p \geqq 4$ and $r = 1$.

5. Some problems where the natural estimator is not minimax

Kiefer [9] and Kudo [10] have shown that under certain conditions, a statistical problem invariant under a group of transformations possesses a minimax solution which is also invariant under this group of transformations. However, these conditions do not hold for the group of all nonsingular linear transformations in a linear space of dimension at least two. I shall give here a problem in multivariate analysis for which I can derive a minimax solution and show that the natural estimator (invariant under the full linear group) is not minimax.

Consider the problem in which we observe $X_1, \cdots, X_n$ independently normally distributed p-dimensional random vectors with mean 0 and unknown covariance matrix Σ where $n \geqq p$. Suppose we want to estimate Σ, say by $\hat{\Sigma}$ with loss function

$$L(\Sigma, \hat{\Sigma}) = \operatorname{tr} \Sigma^{-1} \hat{\Sigma} - \log \det \Sigma^{-1} \hat{\Sigma} - p. \tag{72}$$

The problem is invariant under the transformations $X_i \rightarrow aX_i$, $\Sigma \rightarrow a\Sigma a'$, $\hat{\Sigma} \rightarrow a\hat{\Sigma}a'$ where a is an arbitrary nonsingular $p \times p$ matrix. Also

$$S = \sum_{i=1}^{n} X_i X_i' \tag{73}$$

is a sufficient statistic and if we make the transformation $X_i \rightarrow aX_i$ then $S \rightarrow aSa'$. We may confine our attention to estimators which are functions of S alone. The condition of invariance of an estimator φ (a function on the set of positive definite $p \times p$ symmetric matrices to itself) under transformation by the matrix a is

$$\varphi(asa') = a\varphi(s)a' \qquad \text{for all } s. \tag{74}$$

Let us look for the best estimator φ satisfying (74) for all lower triangular matrices a, that is, those satisfying $a_{ij} = 0$ for $j > i$. We shall find that this $\varphi(S)$ is not a scalar multiple of S. At the end of the section we shall sketch the proof that such an estimator is minimax. Similar results hold for the quadratic loss function

$$L^*(\Sigma, \hat{\Sigma}) = \operatorname{tr} (\Sigma^{-1} \hat{\Sigma} - I)^2 \tag{75}$$

but I have not been able to get an explicit formula for a minimax estimator in this case.

Putting $s = I$ in (74) we find

$$\varphi(aa') = a\varphi(I)a'. \tag{76}$$

When we let a range over the set of diagonal matrices with ± 1 on the diagonal, this yields

$$\varphi(I) = a\varphi(I)a', \tag{77}$$

which implies that $\varphi(I)$ is a diagonal matrix, say Δ, with ith diagonal element Δ_i. This together with (74) determines φ since any positive definite symmetric matrix S can be factored as

$$S = KK' \tag{78}$$

with K lower triangular (with positive diagonal elements) and we then have

$$\varphi(S) = K\Delta K'. \tag{79}$$

Since the group of lower triangular matrices operates transitively on the parameter space, the risk of an invariant procedure φ is constant. Thus we compute the risk only for $\Sigma = I$. We then have

$$\begin{aligned} \rho(I, \varphi) &= E[\operatorname{tr} \varphi(S) - \log \det \varphi(S) - p] \\ &= E(\operatorname{tr} K\Delta K' - \log \det K\Delta K' - p] \\ &= E \operatorname{tr} K\Delta K' - \log \det \Delta - E \log \det S - p. \end{aligned} \tag{80}$$

But

$$\begin{aligned} E \operatorname{tr} K\Delta K' &= \sum_{i,k} \Delta_i EK_{ki}^2 \\ &= \sum \Delta_i E\chi^2_{n-i+1+p-i} = \sum \Delta_i(n + p - 2i + 1) \end{aligned} \tag{81}$$

since the elements of K are independent of each other, the ith diagonal element being distributed as χ_{n-i+1} and the elements below the diagonal normal with mean 0 and variance 1. Also, for the same reason,

$$E \log \det S = \sum_{i=1}^{p} E \log \chi^2_{n-i+1}. \tag{82}$$

It follows that

$$\begin{aligned} \rho(\Sigma, \varphi) &= \rho(I, \varphi) \\ &= \sum_{i=1}^{p} [(n + p - 2i + 1)\Delta_i - \log \Delta_i] - \sum_{i=1}^{p} E \log \chi^2_{n-i+1} - p. \end{aligned} \tag{83}$$

This attains its minimum value of

$$\begin{aligned} \rho(\Sigma, \varphi^*) &= \sum_{i=1}^{p} \left[1 - \log \frac{1}{n + p - 2i + 1} - E \log \chi^2_{n-i+1}\right] - p \\ &= \sum [\log (n + p - 2i + 1) - E \log \chi^2_{n-i+1}] \end{aligned} \tag{84}$$

when

$$\Delta_i = \frac{1}{n + p - 2i + 1}. \tag{85}$$

We have thus found the minimax estimator in a class of estimators which includes the natural estimators (multiples of S) to be different from the natural estimators. Since the group of lower triangular matrices is solvable it follows from the results of Kiefer [9] that the estimator given by (79) and (85) is minimax. However, it is not admissible. One can get a better estimator by averaging this estimator and one obtained by permuting the coordinates, applying the method given above and then undoing the permutation. It must be admitted that the problem is somewhat artificial.

6. Some more unsolved problems

In section 4 several conjectures and unsolved problems concerning estimation of location parameters have been mentioned. Some other problems are listed below. Of course, one can combine these in many ways to produce more difficult problems.

(i) What are the admissible estimators of location parameters? In particular, what are the admissible minimax estimators of location parameters?

(ii) What results can be obtained for more general loss functions invariant under translation?

(iii) For a problem invariant under a group other than a translation group, when is the best invariant estimator admissible? In particular, is Pitman's estimator admissible when both locations and scale parameters are unknown?

(iv) What can we say in the case of more complicated problems where there may be no natural estimator? For example, consider the problem in which we observe $S_1, \cdots, S_n$ independently distributed as $\sigma_n^2 \chi_k^2$, and want to estimate $\sigma_1^2, \cdots, \sigma_n^2$ by $\hat{\sigma}_1^2, \cdots, \hat{\sigma}_n^2$ with loss function

$$L(\sigma_1^2, \cdots, \sigma_n^2; \hat{\sigma}_1^2, \cdots, \hat{\sigma}_n^2) = \frac{\sum (\sigma_i^2 - \hat{\sigma}_i^2)^2}{\sum \sigma_i^4}. \tag{86}$$

It is clear that

$$\hat{\sigma}_i^2 = \frac{1}{k+2} S_i \tag{87}$$

is a minimax estimator since the risk for this estimator is constant and it is minimax when all except one of the σ_i^2 are 0 (see Hodges and Lehmann [7]). But this estimator is clearly very poor if k is small and n is large. This problem arises in the estimation of the covariances in a finite stationary circular Gaussian process.

REFERENCES

[1] A. C. AITKEN, "On least squares and linear combination of observations," *Proc. Roy. Soc. Edinburgh, Sect. A*, Vol. 55 (1935), pp. 42–48.
[2] D. BLACKWELL, "On the translation parameter problem for discrete variables," *Ann. Math. Statist.*, Vol. 22 (1951), pp. 393–399.
[3] C. BLYTH, "On minimax statistical decision procedures and their admissibility," *Ann. Math. Statist.*, Vol. 22 (1951), pp. 22–42.
[4] F. N. DAVID and J. NEYMAN, "Extension of the Markoff theorem on least squares," *Statist. Res. Mem.*, Vol. 1 (1938), pp. 105–116.
[5] R. A. FISHER, "On the mathematical foundations of theoretical statistics," *Philos. Trans. Roy. Soc. London, Ser. A*, Vol. 222 (1922), pp. 309–368.
[6] 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.
[7] 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.
[8] S. KARLIN, "Admissibility for estimation with quadratic loss," *Ann. Math. Statist.*, Vol. 29 (1958), pp. 406–436.
[9] J. KIEFER, "Invariance, minimax sequential estimation, and continuous time processes," *Ann. Math. Statist.*, Vol. 28 (1957), pp. 573–601.
[10] H. KUDO, "On minimax invariant estimators of the transformation parameter," *Nat. Sci. Rep. Ochanomizu Univ.*, Vol. 6 (1955), pp. 31–73.
[11] E. L. LEHMANN, *Testing Statistical Hypotheses*, New York, Wiley, 1959, pp. 231 and 338.
[12] A. MARKOV, *Calculus of Probability*, St. Petersburg, 1908 (2nd ed.). (In Russian.)
[13] E. J. G. PITMAN, "Location and scale parameters," *Biometrika*, Vol. 30 (1939), pp. 391–421.
[14] C. STEIN, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1956, Vol. 1, pp. 197–206.
[15] ———, "A necessary and sufficient condition for admissibility," *Ann. Math. Statist.*, Vol. 26 (1955), pp. 518–522.
[16] ———, "The admissibility of Pitman's estimator for a single location parameter," *Ann. Math. Statist.*, Vol. 30 (1959), pp. 970–979.
[17] ———, "Multiple regression," *Contributions to Probability and Statistics, Essays in Honor of Harold Hotelling*, Stanford, Stanford University Press, 1960, pp. 424–443.
[18] A. WALD, "Contributions to the theory of statistical estimation and testing hypotheses," *Ann. Math. Statist.*, Vol. 10 (1939), pp. 299–326.
[19] R. A. WIJSMAN, "Random orthogonal transformations and their use in some classical distribution problems in multivariate analysis," *Ann. Math. Statist.*, Vol. 28 (1957), pp. 415–423.