COWLES COMMISSION DISCUSSION PAPER: STATISTICS NO. 387

NOTE: Cowles Commission Discussion Papers are preliminary materials circulated privately to stimulate private discussion and are not ready for critical comment or appraisal in publications. References in publications to Discussion Papers (other than mere acknowledgment by a writer that he has had access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

Minimax Estimators in Markoff Situations[1/]

Roy Radner

April 22, 1955

This note considers the problem of minimax estimation in "Markoff situations", for four different assumptions concerning the covariance matrix of the observations. In each case there is some Markoff (minimum variance linear unbiassed) estimator which is minimax among the set of all estimators. In three of the cases the class of covariance matrices within which the true one is assumed to be defined by choosing a particular matrix norm, and requiring that the norm of the covariance matrix be not greater than some given constant; in each of these cases the minimax estimator does not depend upon the actual value of the constant, but may depend upon the form of the norm.

Let y be a random N-dimensional vector with mean and covariance matrix given by:

$$(1)\qquad \begin{aligned} E\,y &= x\,T \\ \mathrm{Cov}\,(y) &= C \end{aligned}$$

1/ Research undertaken by the Cowles Commission for Research in Economics under contract Nonr-358(01), NR 047-006 with the Office of Naval Research.

where T is a known M x N matrix, x is an unknown M-dimensional vector, and C is only known to be in some class $\mathcal{C}$ of N x N covariance matrices. It is required to estimate a given linear function fx' of x, from a single observation on y. The loss incurred in a given instance is taken to be the square of the error. Let $\mathcal{P}$ be any class of distributions P of y, such that any P in $\mathcal{P}$ has property (1) for some x and some C in $\mathcal{C}$, and such that $\mathcal{P}$ contains all normal distributions of y satisfying (1).

The estimator $\hat{\alpha}$ will be called Markoff relative to C if it is the minimum variance linear unbiassed estimator of fx', given C. As is well known, a Markoff estimator is characterized by the fact that it is a linear function $\hat{a}y'$ of y such that $\hat{a}$ minimizes the quadratic form aCa' subject to the constraint aT' = f. The vector $\hat{a}$ depends upon C only up to multiplication by a positive constant (and is in fact equal to $f(TC^{-1}T')^{-1}\, TC^{-1}$, provided the required inverses exist).

The main tool of this note is Lemma 3 below, which is an immediate consequence of the following two lemmas. Let r(a,P) denote the risk (mean squared error) for a given estimator α and a given P in $\mathcal{P}$; and let $R(\alpha,G)$ denote the expected risk for α and an a priori distribution G on $\mathcal{P}$.

<u>Lemma 1</u>. If, for some constant k, $\hat{\alpha}$ satisfies

(a) $r(\hat{\alpha},P) \leq k$ for all P in $\mathcal{P}$,

(b) there is a sequence $\{G_n\}$ of a priori distributions on $\mathcal{P}$ such that

$$\lim_{n \to \infty} \inf_{\alpha} R(\alpha, G_n) = k,$$

then α is minimax.

<u>Lemma 2</u>. If α is Markoff relative to C, then there is seque[illegible] of a priori distributions on the class of normal distributions $N(x$[illegible] $(-\infty < x < \infty)$ of y, such that $\lim_{n \to \infty} \inf_{\alpha} R(\alpha, G_n) = \hat{a}C\hat{a}'$ (where $\hat{\alpha}(y) = \hat{a}y'$).

<u>Proof</u>: Take G_n to be defined by a normal distribution of x, with mean zero and variance n; it can then be verified that this sequence has the desired property. I omit the details.

For any linear unbiassed estimator $\alpha(y) = ay'$, the risk associated with any P in $\mathcal{P}$ depends only on the corresponding C in $\mathcal{C}$, and is in fact equal to aCa'. If in particular, $\hat{\alpha}$ is Markoff relative to C, then the risk $\hat{a}C\hat{a}'$ will be denoted by $\hat{r}(C)$. Lemmas 1 and 2 result immediately in:

<u>Lemma 3</u>. If $\hat{\alpha}$ is Markoff relative to $\hat{C}$ in $\mathcal{C}$, and if $\hat{a}C\hat{a}' \leq \hat{a}\hat{C}\hat{a} = \hat{r}(\hat{C})$ for all C in $\mathcal{C}$, then $\hat{\alpha}$ is minimax among all estimators.

If $\hat{\alpha}$ is minimax, and Markoff relative to $\hat{C}$, as in Lemma 3, then although there may be no least favorable a priori distribution, lemma 2 gives sense to the statement that $\hat{C}$ is a least favorable covariance matrix in $\mathcal{C}$. In all four cases below it is possible to exhibit such a least favorable matrix.

<u>Case I</u>. Let $\mathcal{C}$ be the class of all $C = \sigma^2 C_o$, for which $\sigma^2 \leq k^2$, where C_o is a fixed covariance matrix and k^2 is a fixed constant. Let $\hat{\alpha}$ be the Markoff estimator relative to $k^2 C_o$; then $\hat{\alpha}$ is minimax for $\mathcal{C} = \mathcal{C}_I$. This follows immediately from Lemmas 3 and the fact that for any $C = \sigma^2 C_o$ in $\mathcal{C}_I$, the risk for $\hat{\alpha}$, given C, is

$$\hat{a}\,\sigma^2 C_o \hat{a} \leq k^2 \hat{a}\, C_o\, \hat{a} = \hat{r}\,(k^2 C_o)$$

where $\hat{\alpha}(y) = \hat{a}y$, does not depend upon the value of k^2.

The assumption of Case I is quite restrictive. On the other hand, if nothing at all is known about the covariance matrix, i.e. if $\mathcal{C}$ is the class of all $N \times N$ covariance matrices, then the risk for any estimator is unbounded. Cases II - IV indicate the effects of restricting $\mathcal{C}$ in various ways.

Case II. Let $\mathcal{C}_{II}$ be the class of all C for which $||C|| \leq k^2$, where k^2 is fixed, and

$$||C|| = \max_{||a||=1} a\,C\,a'$$

This says that the variance of any linear combination ay' for which $\Sigma a_i^2 = 1$ is not greater than k^2. ($||C||$ is also the largest characteristic root of C.) Let $\hat{\alpha}$ be Markoff relative to $k^2 I$; then $\hat{\alpha}$ is minimax when $\mathcal{C} = \mathcal{C}_{II}$, since for any C in $\mathcal{C}_{II}$:

$$\hat{\alpha}\, C\, \alpha \leq ||C||\ ||\hat{\alpha}||^2 \leq k^2\, ||\hat{\alpha}||^2 = \hat{r}\,(k^2 I)$$

and Lemma 3 applies.

Note that the least favorable C is one in which the coordinates of y are uncorrelated, and that $\hat{\alpha}$ does not depend upon the value of k^2.

Case III. Let $\mathcal{C}_{III}$ be the class of all C for which $Var(y_i) = c_{ii} \leq k_i^2$, where the $k_i > 0$ are fixed. Let $\hat{a}$ minimize $(\sum_i k_i\, |a_i|)^2$ subject to $aT' = f$, and let

$$r_i = \begin{cases} \dfrac{\hat{a}_i}{|\hat{a}_i|}, & \text{if } \hat{a}_i \neq 0 \\ 1, & \text{otherwise} \end{cases}$$

Then the matrix $\hat{C} = ((k_i r_i k_j r_j))$ is a covariance matrix of rank 1, and

$$\hat{a}\, \hat{C}\, \hat{a}' = (\sum_i k_i\, |\hat{a}_i|)^2 = \hat{r}\,(\hat{C}).$$

The set of a for which $\sum k_i\, |a_i| \leq k$, is, for any k, the convex polyhedron bounded by the 2^N linear varieties $\sum c_i k_i a_i = k$, where the c_i are taken to be ± 1 in all possible ways. The vertices of this polyhedron are the vectors $\pm \frac{k}{k_i} e_i$, where e_i is the i'th unit vector; hence $\hat{a}$ may always be taken to be one of the vectors $\pm \frac{r(\hat{C})}{k_i} e_i$.

For any C in $\mathcal{C}_{III}$ for which $c_{ii} > 0$,

$$\hat{a}\, C\, \hat{a}' = \sum_{i,j} \frac{c_{ij}}{\sqrt{c_{ii} c_{jj}}} \hat{a}_i \sqrt{c_{ii}}\ \hat{a}_j \sqrt{c_{jj}}$$

$$\leq \sum_{ij} |\hat{a}_i| \sqrt{c_{ii}}\ |\hat{a}_j| \sqrt{c_{jj}}$$

$$= \left(\sum_i |\hat{a}_i| \sqrt{c_{ii}} \right)^2$$

$$\leq \left(\sum_i |a_i|\, k_i \right)^2 = \hat{r}\,(\hat{C})$$

so that Lemma 3 applies and $\hat{\alpha}\,(y) = \hat{a}\, y'$ is minimax for $\mathcal{C}_{III}$. In this case, the least favorable C is $\hat{C}$, i.e., the coordinates of y are perfectly correlated. The estimator $\hat{\alpha}$ depends upon the k_i only up to a constant of proportionality.

In particular, if all the $k_i = k$, $\hat{\alpha}$ does not depend upon k. In this last case of equal k_i, $\mathcal{C}_{III}$ could also be defined by: $|c_{ij}| \leq k^2$.

<u>Case IV.</u> Let $\mathcal{C}_{IV}$ be the class of all C for which trace $(C) = \sum_i c_{ii} \leq k^2$. Let $\hat{\alpha}$ be Markoff relative to $k^2 I$. For any C in $\mathcal{C}_{IV}$, let λ_i be its characteristic roots and p_i the corresponding characteristic vectors; then $\lambda_i \geq 0$, $\sum \lambda_i \leq k$, and

$$\hat{a}\, C\, \hat{a} = \sum \lambda_i\, (a, p_i)^2 \leq \sum \lambda_i\, \hat{a}\, \hat{a}' \leq k^2\, \hat{a}\, \hat{a}' = \hat{r}\,(k^2\, I),$$

so that $\hat{\alpha}$ is minimax for $\mathcal{C}_{IV}$, giving the same result as Case I.