ALL ADMISSIBLE LINEAR ESTIMATES OF THE MEAN VECTOR[1]

By Arthur Cohen

Rutgers—The State University

1. Introduction and summary. Let y be a single observation on a $p \times 1$ random vector which is distributed according to the multivariate normal distribution with mean vector θ and covariance matrix I. Consider the problem of estimating θ when the loss function is the sum of the squared errors in estimating the individual components of θ. Let G be a $p \times p$ real matrix. Then we will prove that the estimate Gy is admissible if and only if G is symmetric, and the characteristic roots of G, g_i say, $i = 1, 2, \cdots, p$, satisfy $0 \leqq g_i \leqq 1$, with equality at one for at most two of the roots.

The proof concerning the characteristic roots uses the results of Karlin [4] and James and Stein [3]. We give two proofs of the fact that Gy is inadmissible when G is asymmetrical. In the first proof we give an estimate G^*y that is better than Gy, where G^* is a symmetric matrix. This not only adds to the practicality of the result, but also enables us to resolve the question of which estimates are admissible in the restricted class of estimates of the form Gy. The method of the second proof, which utilizes a theorem of Sacks [6], leads to the following finding: If Gy is admissible, then Gy must be a generalized Bayes procedure, where the unique generalized prior distribution must be either a multivariate normal distribution with mean vector zero and a specified covariance matrix determined by G, or the product of a distribution which is multivariate normal over a subspace of the parameter space and a distribution which is uniformly distributed over a subspace of the parameter space. This latter finding generalizes the well known one dimensional case.

In Section 2 then the main results are proved. In Section 3 some remarks concerned with generalizing the main results are given. We remark here that the decision theory terminology used is more or less that of Blackwell and Girshick [1].

2. Main results. In this section we prove the following:

Theorem 2.1. *The estimate Gy is admissible if and only if G is symmetric, and the characteristic roots of G, say g_i, satisfy, $0 \leqq g_i \leqq 1$, with equality at one for at most two of the roots.*

Proof. The proof will be given in two parts. In the first part we restrict ourselves to symmetric matrices. In the second part we show Gy is inadmissible when G is asymmetrical.

To prove the result for symmetric matrices we need the following:

Lemma 2.1. *Let G be any matrix and let P be any orthogonal matrix. Then the estimate $(P'GP)y$ is admissible if and only if Gy is admissible.*

Received 7 May 1965; revised 25 October 1965.

[1] Research was supported by the National Science Foundation under Grant GP-4644.

PROOF. Let $L(\psi(\theta), \delta(y))$ be the loss for estimating $\psi(\theta)$ by $\delta(y)$. Then we write the risk as

$$\rho(\psi, \delta; \theta) = E_\theta L(\psi(\theta), \delta(y)).$$

Now if P is orthogonal, if ψ_1, δ_1, are given, and if ψ_2, δ_2 are defined by $\psi_2(\theta) = \psi_1(P\theta)$, $\delta_2(y) = \delta_1(Py)$ then

$$(2.1) \qquad \begin{aligned} \rho(\psi_2, \delta_2; \theta) &= E_\theta L(\psi_1(P\theta), \delta_1(Py)) \\ &= E_{P\theta} L(\psi_1(P\theta), \delta_1(y)) = \rho(\psi_1, \delta_1; P(\theta)), \end{aligned}$$

where the next to the last equality in (2.1) follows from the fact that Py under θ has the same distribution as y under $P\theta$. Thus we have $\rho(\psi_2, \delta_2; \theta) = \rho(\psi_1, \delta_1; P\theta)$, and from this follows that δ_2 is admissible for ψ_2 if and only if δ_1 is admissible for ψ_1. If we set $\psi_1(\theta) = \theta$ and $\delta_1(y) = Gy$ then we get GPy is admissible for $P\theta$ if and only if Gy is admissible for θ. Also, since for our particular loss function, we get $L(P\psi, P\delta) = L(\psi, \delta)$ it follows that $P'GPy$ is admissible for estimating θ if and only if GPy is admissible for estimating $P\theta$. Thus $P'GPy$ is admissible for estimating θ if and only if Gy is admissible for estimating θ. This completes the proof of the lemma.

We now return to the proof of the theorem for symmetric matrices. Since every symmetric matrix can be written as $P'DP$, for P an orthogonal matrix and D a diagonal matrix, it follows from Lemma 2.1, where we now let G be a diagonal matrix, that we need only prove the desired results for diagonal matrices. Hence consider a diagonal matrix D with diagonal elements d_i. Then the estimate Dy is admissible for $0 \leqq d_i < 1$, $i = 1, 2, \cdots, p$, since Dy is the unique Bayes solution with respect to the prior distribution

$$d\xi(\theta) = (2\pi)^{-p/2} \prod_{i=1}^p \lambda_i^{\frac{1}{2}} e^{-\lambda_i \theta_i^2/2} \prod_{i=1}^p d\theta_i,$$

where $\lambda_i = (1 - d_i)/d_i$. (Of course if any d_i is zero, then the marginal prior distribution for θ_i puts all its probability at $\theta_i = 0$.) Now if any d_i is negative it is easy to verify that replacing d_i by $-d_i$ leads to a better estimate. If any $d_i > 1$, it is easy to verify that replacing d_i by 1 leads to a better estimate. Also if three or more $d_i = 1$, while all other d_i satisfy, $0 \leqq d_i < 1$, it follows immediately from Stein [3], that by replacing those $d_i = 1$, by $(1 - (k-2)/\sum_{i=1}^p y_i^2)$, where k is the number of the d_i that equal 1, we are led to a better estimate. Hence for diagonal matrices, it only remains to prove that Dy is admissible if one or two $d_i = 1$, while for all others, $0 \leqq d_i < 1$. Let $d_1 = d_2 = 1$, and $0 \leqq d_i < 1$, for $i = 3, 4, \cdots, p$, since the other cases can be treated similarly. Now this fact also follows from Stein [3]. To see this suppose such a Dy is inadmissible. Then there exists an estimate $H'(y) = (h_1(y), h_2(y), \cdots, h_p(y))$, such that $H(y)$ is better than Dy. That is,

$$(2.2) \qquad \rho(\theta, H(y)) \leqq \rho(\theta, Dy),$$

for every θ with strict inequality for at least some θ. If we evaluate the risk for Dy, we find that (2.2) can be written as

$$\sum_{i=1}^{p} \int_{E_p} (h_i(y) - \theta_i)^2 \, d\Phi(y; \theta, I) \leqq 2 + \sum_{i=3}^{p} (d_i^2 + \theta_i^2(d_i - 1)^2), \tag{2.3}$$

where E_p is the p dimensional sample space and $\Phi(x; \nu, \Sigma)$ is the probability distribution function of a $p \times 1$ random vector x, distributed as multivariate normal with mean vector ν and covariance matrix Σ. Now if we multiply both sides of (2.3) by $\prod_{i=3}^{p} (\lambda_i/2\pi)^{\frac{1}{2}} e^{-\lambda_i \theta_i^2/2} = \varphi(\theta)$, and integrate with respect to θ_i, $i = 3, 4, \cdots, p$, over the entire range of these θ_i, we get

$$\begin{aligned} &\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{E_p} [(h_1(y) - \theta_1)^2 + (h_2(y) - \theta_2)^2] \, d\Phi(y; \theta, I)\varphi(\theta) \prod_{i=3}^{p} d\theta_i \\ &\quad + \sum_{j=3}^{p} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{E_p} (h_j - \theta_j)^2 \, d\Phi(y; \theta, I)\varphi(\theta) \\ &\quad \cdot \prod_{i=3}^{p} d\theta_i \leqq 2 + \sum_{i=3}^{p} d_i . \end{aligned} \tag{2.4}$$

But $\sum_{i=3}^{p} d_i$ is the minimum value attainable for the second term on the left-hand side of (2.4) and so after some integration we are led from (2.4) to

$$\int_{E_p} [(h_1(y) - \theta_1)^2 + (h_2(y) - \theta_2)^2] \, d\Phi(y; \mu, \Sigma) \leqq 2, \tag{2.5}$$

where $\mu' = (\theta_1, \theta_2, 0, \cdots, 0)$ and Σ is a diagonal matrix with elements $\sigma_{11} = 1$, $\sigma_{22} = 1$, $\sigma_{ii} = (\lambda_i + 1)/\lambda_i$, $i = 3, \cdots, p$. It is clear from (2.5) and sufficiency then, that the estimates $h_1(y)$ and $h_2(y)$ need only be functions of y_1 and y_2. Hence it now follows from Stein [3], that $h_1(y) = y_1$, $h_2(y) = y_2$, and $h_i(y) = d_i y_i$, for $i = 3, 4, \cdots, p$. This proves that Dy is admissible and finishes the proof of the first part of the theorem.

We now prove the second part of the theorem. That is, for any estimate Gy, where G is asymmetrical, there exists a better estimate. In fact, given G asymmetrical, G^*y is better where $G^* = I - [(G - I)'(G - I)]^{\frac{1}{2}}$. For note that the risk for any estimate Gy may be written as

$$\rho(\theta, Gy) = E(Gy - \theta)'(Gy - \theta) = \operatorname{tr}(G'G) + \theta'(G - I)'(G - I)\theta.$$

Therefore we need only show that

$$\operatorname{tr}(G'G) > \operatorname{tr}(G^{*\prime}G^*). \tag{2.6}$$

But since $(G^* - I)'(G^* - I) = (G - I)'(G - I)$, it follows that (2.6) is equivalent to $\operatorname{tr}(G - I) > \operatorname{tr}(G^* - I) = -\operatorname{tr}\{[(G - I)'(G - I)]^{\frac{1}{2}}\}$, or to

$$\operatorname{tr}\{[(I - G)'(I - G)]^{\frac{1}{2}}\} > \operatorname{tr}(I - G). \tag{2.7}$$

Now for any real matrix, A say, $\operatorname{tr}[(A'A)^{\frac{1}{2}}] \geqq \operatorname{tr} A$ (see, for example [5], Section 4.2). Furthermore, it can be verified that if A is asymmetric then $\operatorname{tr}[(A'A)^{\frac{1}{2}}] > \operatorname{tr} A$. Hence (2.7) is true and the proof of the second part of the theorem is complete.

A second proof of the fact that Gy is inadmissible for G asymmetrical is as follows:

Suppose G is not symmetric and Gy is admissible. Then it follows from the generalization of a theorem due to Sacks [6], Remark 4, p. 767, that the ith component of Gy can be written as

$$(2.8) \quad \sum_{k=1}^{p} g_{ik}y_k = \int_{\Omega_p} \theta_i[\prod_{j=1}^{p} \beta(\theta_j)e^{y_j\theta_j}]\, d\xi(\theta)/\int_{\Omega_p} [\prod_{j=1}^{p} \beta(\theta_j)e^{y_j\theta_j}]\, d\xi(\theta),$$

where g_{ik} is the (ik)th element of G, $\beta(\theta_j) = e^{-\theta_j^2/2}$, Ω_p is the p dimensional parameter space and $\xi(\theta) = \xi(\theta_1, \theta_2, \cdots, \theta_p)$, is some generalized prior distribution. We now demonstrate that the (rs)th element of G must be equal to the (sr)th element of G. For, if we differentiate the rth element of Gy with respect to y_s, we get from (2.8),

$$(2.9) \quad \begin{aligned} g_{rs} = {} & \int_{\Omega_p} [\prod_{j=1}^{p} \beta(\theta_j)e^{y_j\theta_j}]\, d\xi(\theta) \int_{\Omega_p} \theta_r\theta_s[\prod_{j=1}^{p} \beta(\theta_j)e^{y_j\theta_j}]\, d\xi(\theta)/ \\ & (\int_{\Omega_p} [\prod_{j=1}^{p} \beta(\theta_j)e^{y_j\theta_j}]\, d\xi(\theta))^2 \\ & - \int_{\Omega_p} \theta_r[\prod_{j=1}^{p} \beta(\theta_j)e^{y_j\theta_j}]\, d\xi(\theta)\int_{\Omega_p} \theta_s[\prod_{j=1}^{p} \beta(\theta_j)e^{y_j\theta_j}]\, d\xi(\theta)/ \\ & (\int_{\Omega_p} [\prod_{j=1}^{p} \beta(\theta_j)e^{y_j\theta_j}]\, d\xi(\theta))^2. \end{aligned}$$

Similarly, if we differentiate the sth element of Gy with respect to y_r, we get g_{sr} equal to the right-hand side of (2.9). Hence $g_{rs} = g_{sr}$ and the matrix G is symmetric. This is a contradiction, from which we may conclude that Gy, with G asymmetric, is inadmissible.

The latter proof of the inadmissibility of Gy, for G asymmetrical suggests the following:

THEOREM 2.2. *If Gy is admissible, then Gy must be a generalized Bayes procedure. The generalized prior distribution must be unique and be either a multivariate normal distribution with mean vector* 0 *and covariance matrix* Σ, *where* Σ *is determined by the relation* $G = \Sigma(\Sigma + I)^{-1}$ *or the prior distribution must be a product of a distribution which is multivariate normal over a subspace of the parameter space and a distribution which is uniformly distributed over a subspace of the parameter space.*

PROOF. It follows immediately from Sacks [6] that if Gy is admissible it must be a generalized Bayes procedure. Therefore the ith component of Gy can be written as in (2.8), which in turn can be written as:

$$(2.10) \quad \sum_{k=1}^{p} g_{ik}y_k = (\partial/\partial y_i) \log_e \int_{\Omega_p} [\prod_{j=1}^{p} \beta(\theta_j)e^{y_j\theta_j}]\, d\xi(\theta).$$

If we now take the indefinite integral of both sides of (2.10), and then consider each side as the exponent of the constant e, we get for each $i = 1, 2, \cdots, p$,

$$(2.11) \quad \begin{aligned} \exp \{g_{ii}y_i^2/2 + \textstyle\sum_{k=1,k\neq i}^{p} g_{ik}y_iy_k + C_i(y_1, y_2, \cdots, y_{i-1}, y_{i+1}, \cdots, y_p)\} \\ = \int_{\Omega_p} [\prod_{j=1}^{p} \beta(\theta_j)e^{y_j\theta_j}]\, d\xi(\theta), \end{aligned}$$

where $C_i(y_1, y_2, \cdots, y_{i-1}, y_{i+1}, \cdots, y_p)$ is constant as a function of y_i, but may be a function of $y_1, y_2, \cdots, y_{i-1}, y_{i+1}, \cdots, y_p$. Since (2.11) is true for each $i = 1, 2, \cdots, p$, we may determine the functions C_i from the p equations. That is, the set (2.11) yields

$$(2.12) \quad \begin{aligned} & g_{ii}y_i^2/2 + \textstyle\sum_{k=1,k\neq i}^{p} g_{ik}y_iy_k + C_i(y_1, y_2, \cdots, y_{i-1}, y_{i+1}, \cdots, y_p) \\ & = g_{jj}y_j^2/2 + \textstyle\sum_{k=1,k\neq j}^{p} g_{jk}y_jy_k + C_j(y_1, y_2, \cdots, y_{j-1}, y_{j+1}, \cdots, y_p), \\ & \qquad \text{for} \quad i \neq j,\ i, j = 1, 2, \cdots, p. \end{aligned}$$

If we use the fact established earlier that G must be symmetric, we find that the Equations (2.12) imply that

$$C_i(y_1, y_2, \cdots, y_{i-1}, y_{i+1}, \cdots, y_p) = \sum_{k=1, k\neq i}^{p} g_{kk} y_k^2/2 + \sum_{r<s; r,s\neq i; r,s=1}^{p} g_{rs} y_r y_s + c,$$

for some constant c. This, along with (2.11) imply that we have

$$\exp\{\tfrac{1}{2} y' G y\} = \int_{\Omega_p} [\prod_{j=1}^{p} \beta(\theta_j) e^{y_j \theta_j}] \, d\xi(\theta). \tag{2.13}$$

We now recognize that the right-hand side of (2.13) may be regarded as a moment generating function. Since the left-hand side may be regarded as the moment generating function of a multivariate normal distribution (even if G is a singular matrix), it follows that $\xi(\theta)$ is uniquely determined. Also if all the characteristic roots of G satisfy, $0 \leqq g_i < 1$, then $\xi(\theta)$ is multivariate normal with mean vector 0 and covariance matrix Σ, where $G = \Sigma(\Sigma + I)^{-1}$. (See [2], Lemma 2.2.) If one or two of the $g_i = 1$ (say $g_{p-1} = 1, g_p = 1$, since the other cases can be treated similarly), then the prior distribution for θ is determined from G as follows: Let $z = Py$ and $\mu = P\theta$, where P is an orthogonal matrix such that $G = P'DP$, and D is diagonal with elements g_i. We define the prior distribution of μ which in turn determines the prior distribution for θ. Let then the prior distribution for μ be

$$d\xi(\mu) = (2\pi)^{-(p-2)/2} \prod_{i=1}^{p-2} \lambda_i^{\frac{1}{2}} e^{-\lambda_i \mu_i^2/2} \prod_{i=1}^{p} d\mu_i,$$

where $\lambda_i = (1 - g_i)/g_i$. Then the generalized Bayes solution with respect to this generalized prior distribution is $E(\theta/y) = E(P'\mu/y) = P'E(\mu/y) = P'E(\mu/z) = P'Dz = Gy$. Thus we have shown the relationship between G and the prior distribution for which G is a generalized Bayes procedure. This completes the proof of Theorem 2.2.

3. Concluding remarks and generalizations. In this section we give some remarks which indicate generalizations of the main results.

1. There is no loss in generality in assuming y to be a single observation. If $y_1, y_2, \cdots, y_n$ were a random sample, then the development, applying to $\bar{y}$, the sample mean, would follow.

2. The development also follows if y is multivariate normal with mean vector θ and covariance matrix $\sigma^2 I$, where σ^2 may be unknown. Some additional conditions, as given in Stein [3], p. 366, are required for some of the boundary cases.

3. If we restrict the class of estimates to be linear, that is, of the form Gy, then Theorem 2.1 is true with the revision that all the characteristic roots of G may be equal to one.

4. Also if we restrict the class of estimates to be linear, and relax the normality assumption, the revision of Theorem 2.1, in Remark 3, is correct. This follows since the risk function for such a problem depends only on second moments. It is also correct if y has covariance matrix $\sigma^2 I$, with σ^2 unknown.

5. For non-homogeneous linear estimates, that is estimates of the form $Gy + k$,

for k a vector of constants, if all $0 \leqq g_i < 1$, and G is symmetric then all such estimates are admissible. If some of the g_i equal to one, say g_{p-1} and g_p, then in order for $Gy + k$ to be admissible, the vector k would have to be such that the last two components of Pk would be zero, where P is the orthogonal matrix such that $G = P'DP$, with D diagonal.

I would like to thank Professor Charles Stein and Dr. Colin Mallows for helpful remarks.

REFERENCES

[1] Blackwell, D. and Girshick, M. A. (1954). *Theory of Games and Statistical Decisions*. Wiley, New York.
[2] Cohen, Arthur (1965). Estimates of linear combinations of the parameters in the mean vector of a multivariate distribution. *Ann. Math. Statist.* **36** 78–87.
[3] James, W. and Stein, Charles (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 361–379. Univ. of California Press.
[4] Karlin, Samuel (1958). Admissibility for estimation with quadratic loss. *Ann. Math. Statist.* **29** 406–436.
[5] Marcus, M. and Minc, H. (1964). *A Survey of Matrix Theory and Matrix Inequalities*. Allyn and Bacon, Boston.
[6] Sacks, Jerome (1963). Generalized Bayes solutions in estimation problems. *Ann. Math. Statist.* **34** 751–768.