The Annals of Statistics
1987, Vol. 15, No. 4, 1639–1650

BEST EQUIVARIANT ESTIMATORS OF A CHOLESKY DECOMPOSITION[1]

By Morris L. Eaton and Ingram Olkin

University of Minnesota and Stanford University

Every positive definite matrix Σ has a unique Cholesky decomposition $\Sigma = \theta\theta'$, where θ is lower triangular with positive diagonal elements. Suppose that S has a Wishart distribution with mean $n\Sigma$ and that S has the Cholesky decomposition $S = XX'$. We show, for a variety of loss functions, that XD, where D is diagonal, is a best equivariant estimator of θ. Explicit expressions for D are provided.

1. Introduction. Given a $p \times p$ positive definite matrix A, there exists a unique $p \times p$ lower triangular matrix T with positive diagonal elements such that

$$A = TT'. \tag{1.1}$$

This is commonly called the Cholesky decomposition of A and is a basic decomposition in numerical analysis because it enables one to solve systems of linear equations more easily.

Suppose that a $p \times p$ random matrix S has a Wishart distribution $\mathscr{W}(\Sigma, p, n)$, where Σ is a $p \times p$ positive definite matrix, and the degrees of freedom n is at least p, so that S is positive definite with probability 1. It is well known that

$$\hat{\Sigma} \equiv \frac{1}{n} S \tag{1.2}$$

is an unbiased estimator of Σ. Now, write Σ in its Cholesky decomposition

$$\Sigma = \theta\theta', \tag{1.3}$$

where θ is a $p \times p$ lower triangular matrix with positive diagonal elements. The problem we discuss here is the estimation of the matrix θ based on the data S.

An unbiased estimator of θ was obtained by Olkin (1985). [See Lehmann (1983), page 85, for the one dimensional case.] We describe this estimator since it motivates the remainder of this paper. Write

$$S = XX', \tag{1.4}$$

where X is lower triangular with positive diagonal elements. In the statistical literature, the x_{ij} are called rectangular coordinates and the factorization is sometimes called the Bartlett decomposition. The joint distribution of the x_{ij} was obtained by Bartlett (1933) and by Mahalanobis, Bose and Roy (1937). Then

Received July 1986; revised February 1987.

[1] Work supported in part by the National Science Foundation.

AMS 1980 *subject classifications*. 62H10, 15A52, 15A23.

Key words and phrases. Rectangular coordinates, random matrices.

the unique unbiased estimator of θ is

(1.5) $$\hat{\theta} = XD^{-1},$$

where D is a known diagonal matrix. To see how D arises, recall that the density of X is known and can be derived from the Wishart density. Let G^+ be the set of all $p \times p$ lower triangular matrices with positive diagonal elements. Then G^+ is a group and a left Haar measure on G^+ is

(1.6) $$\nu(dx) = \left(\prod_{i=1}^{p} x_{ii}^{i}\right)^{-1} dx,$$

where

$$x = \begin{pmatrix} x_{11} & 0 & & \cdots & 0 \\ x_{21} & x_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \\ x_{p1} & x_{p2} & \cdot & \cdot & x_{pp} \end{pmatrix}$$

and dx is Lebesgue measure on G^+. The density of X with respect to ν is

(1.7) $$p(x|\theta) = \phi(\theta^{-1}x),$$

where

(1.8) $$\phi(u) = k(\det u)^n \exp\left[-\tfrac{1}{2}\operatorname{tr} uu'\right], \qquad u \in G^+,$$

and k is a known constant. Thus, we see that the random matrix

(1.9) $$U = \theta^{-1}X$$

has a density on G^+ given by ϕ in (1.8).

Now, it is known that the elements u_{ij} of U are independently distributed with

(1.10) $$\mathscr{L}(u_{ij}) = \mathscr{N}(0,1), \qquad i > j,$$
$$\mathscr{L}(u_{ii}^2) = \chi^2_{n-i+1}, \qquad i = 1, \ldots, p.$$

Thus

(1.11) $$\mathscr{E}U = D,$$

where D is a diagonal matrix with ith-diagonal element equal to

(1.12) $$\mathscr{E}u_{ii} = \mathscr{E}\left(\chi^2_{n-i+1}\right)^{1/2} = \sqrt{2}\,\Gamma\left(\frac{n-i+2}{2}\right)\Big/\Gamma\left(\frac{n-i+1}{2}\right).$$

The notation in (1.11) is vector space notation for expectation. Thus, (1.11) means that the expectation of any element of U is the corresponding element of D.

In (1.9), note that $\theta^{-1}X$ can be thought of as a linear transformation (determined by θ^{-1}) evaluated at X. Since expectation of vectors commutes with linear transformations, (1.9) and (1.11) yield

(1.13) $$\mathscr{E}XD^{-1} = \mathscr{E}\theta\theta^{-1}XD^{-1} = \theta(\mathscr{E}\theta^{-1}X)D^{-1} = \theta.$$

Hence, XD^{-1} is an unbiased estimator of θ.

Some important observations regarding the preceding argument are:

1. The sample space and parameter space are both the group G^+. The family of distributions of X is invariant under the group action defined by G^+.
2. The random group element X can also be thought of as a random vector. The group action $X \to gX$ defines a linear transformation (determined by g) acting on the vector X.

These two observations are exploited in this paper, where we examine other estimators of θ derived from decision theoretic considerations.

2. The estimation problem. Fix a locally compact group G and let ν denote a left Haar measure on G. Consider a fixed density (on G) with respect to ν, say ϕ. Then, a parametric family of densities on G can be defined by

$$p(x|\theta) = \phi(\theta^{-1}x), \qquad \theta, x \in G. \tag{2.1}$$

Given an X with density (2.1), the sample space and parameter space are G. The problem considered is the estimation of θ based on X. The performance of an estimator is measured by an invariant loss function $L(a, \theta)$, $a, \theta \in G$. That is, L satisfies

$$L(a, \theta) = L(ga, g\theta), \qquad a, \theta, g \in G. \tag{2.2}$$

Choosing $g = \theta^{-1}$ in (2.2) shows that L is invariant if and only if L has the form

$$L(a, \theta) = K(\theta^{-1}a), \tag{2.3}$$

where K is an arbitrary function defined on G.

Because the estimation problem under study is invariant, attention is restricted to equivariant estimators. If t: $G \to G$ is an estimator (a point estimator), the equivariance of t means that

$$t(gx) = gt(x), \tag{2.4}$$

which implies (by choosing $g = x^{-1}$) that

$$t(x) = xt(e), \tag{2.5}$$

where e is the identity in the group. Setting $a = t(e)$, we see that t is equivariant if and only if

$$t(x) = xa \tag{2.6}$$

for some $a \in G$. Thus, choosing an equivariant estimator is equivalent to choosing $a \in G$ and using the estimator in (2.6). Estimators are compared in terms of the risk function. For any equivariant estimator t,

$$\begin{aligned} R(t, \theta) &= \int L(t(x), \theta) p(x|\theta) \nu(dx) = \int K(\theta^{-1}t(x)) \phi(\theta^{-1}x) \nu(dx) \\ &= \int K(\theta^{-1}xa) \phi(\theta^{-1}x) \nu(dx) = \int K(xa) \phi(x) \nu(dx) \\ &= \int K(xa) p(x|e) \nu(dx) \\ &= \psi(a). \end{aligned}$$

Thus, the best equivariant estimator is found by minimizing

$$\psi(a) = \mathscr{E}_e K(Xa), \tag{2.7}$$

where $\mathscr{E}_e$ is the expectation computed at $\theta = e$, the identity in G: In the case that $G = G^+$, these estimators will be minimax because the group G^+ is solvable [Kiefer (1957)].

Some minimization results. As noted in (2.7), the best equivariant estimator is found as the result of a minimization. We first gather several minimization results needed in later examples. The notation $A > 0$ means that A is positive definite. For $A > 0$, the matrix $A^{1/2}$ is the unique positive definite square root. All matrices in this section are $p \times p$.

LEMMA 2.1. *If* $\Delta > 0$, *then*

$$\min_{B>0} [\operatorname{tr} \Delta B - \log \det B] = p + \log \det \Delta$$

is achieved at $B = \Delta^{-1}$.

LEMMA 2.2. *If* $\Delta > 0$ *and* $\Lambda > 0$, *then*

$$\min_{B>0} [\operatorname{tr} \Delta B + \operatorname{tr} \Lambda B^{-1}] = 2 \operatorname{tr}(\Lambda^{1/2} \Delta \Lambda^{1/2})^{1/2}$$

is achieved at $B = \Lambda^{1/2}(\Lambda^{1/2}\Delta\Lambda^{1/2})^{-1/2}\Lambda^{1/2}$.

LEMMA 2.3. *If* $C(Y)$ *is a positive definite matrix that is a function of* Y, *then*

$$\begin{aligned}\min_a \mathscr{E} \operatorname{tr}(a - Y)(a - Y)'C(Y) &= \min_a \operatorname{tr}[aa'\mathscr{E}C(Y) - 2a\mathscr{E}Y'C(Y) + \mathscr{E}YY'C(Y)] \\ &= \operatorname{tr}\left[\mathscr{E}(Y'C(Y)Y) - (\mathscr{E}Y'C(Y))(\mathscr{E}C(Y))^{-1}(\mathscr{E}C(Y)Y)\right],\end{aligned}$$

where the minimum is over all $p \times p$ *matrices* a, *is achieved at* $a = (\mathscr{E}C(Y))^{-1}(\mathscr{E}C(Y)Y)$.

PROOFS OF THE LEMMAS. Each of the functions to be minimized is strictly convex. In Lemma 2.1, $\operatorname{tr} \Delta B$ is linear and the determinant of a positive definite matrix is log concave [e.g., see Marshall and Olkin (1979), page 476]. In Lemma 2.2, the inverse of a positive definite matrix is convex so that $\operatorname{tr} \Lambda B^{-1}$ is convex [see Marshall and Olkin (1979), page 469]. Finally, $(a - Y)'C(Y)(a - Y)$ is convex for each fixed Y and, hence, is convex after averaging.

Since $(\operatorname{tr} \Delta B - \log \det B) \to \infty$ and $(\operatorname{tr} \Delta B + \operatorname{tr} \Lambda B^{-1}) \to \infty$ as B approaches the boundary, we need only examine the first derivative equations. In Lemma 2.1, the derivative equation is $\Delta - B^{-1} = 0$, which immediately yields the result. In Lemma 2.2, we obtain $\Delta - B^{-1}\Lambda B^{-1} = 0$, which is quadratic and readily solved. In Lemma 2.3, we obtain $2(\mathscr{E}C(Y))a - 2\mathscr{E}C(Y)Y = 0$, which yields the result. □

Lemma 2.4. *Let L_1, L_2, L_3 and L_4 denote diagonal matrices with positive diagonals, partitioned $L_i = \operatorname{diag}(l_{i1}, l_{i2})$, $i = 1, \dots, 4$. The minimizer a_0 of*

$$\psi(a) = \operatorname{tr}(a - L_1)L_2(a - L_1)' + \operatorname{tr}(a^{-1} - L_3)L_4(a^{-1} - L_3)'$$

over the set of lower triangular matrices occurs at a_0 diagonal.

Proof. Partition

$$a = \begin{pmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{pmatrix}, \qquad a^{-1} = \begin{pmatrix} a_{11}^{-1} & 0 \\ -a_{22}^{-1}a_{21}a_{11}^{-1} & a_{22}^{-1} \end{pmatrix},$$

where a_{11} is a scalar and a_{22} is a lower triangular matrix of order $p - 1$. Then

$$\begin{aligned}\psi(a) = {} & \operatorname{tr}(a_{11} - l_{11})^2 l_{21} + \operatorname{tr} a_{21}l_{21}a_{21}' + \operatorname{tr}(a_{22} - l_{12})l_{22}(a_{22} - l_{12})' \\ & + \operatorname{tr}\left(a_{11}^{-1} - l_{31}\right)^2 l_{41} + \operatorname{tr} a_{22}^{-1}a_{21}a_{11}^{-1}l_{41}a_{11}^{-1}a_{21}'a_{22}'^{-1} \\ & + \operatorname{tr}\left(a_{22}^{-1} - l_{32}\right)l_{42}\left(a_{22}^{-1} - l_{32}\right)'.\end{aligned}$$

The terms involving a_{21} are

$$\operatorname{tr} a_{21}l_{21}a_{21}' + \operatorname{tr} a_{22}^{-1}a_{21}a_{11}^{-1}l_{41}a_{11}^{-1}a_{21}'a_{22}'^{-1} \geq 0,$$

with equality when $a_{21} = 0$.

An iterative argument shows that every nondiagonal element of the minimizer a_0 must be zero. □

From Lemma 2.4, the minimizer of $\psi(a)$ is a_0 diagonal, the elements of which can be obtained from p separate univariate minimization problems, each of the form: Minimize over $z > 0$,

$$\xi(z) = (z - q_1)^2 q_2 + \left(\frac{1}{z} - q_3\right)^2 q_4, \tag{2.8}$$

where $q_i > 0$, $i = 1, \dots, 4$. Since $\xi(z) > 0$, $\xi(0) = \infty$ and $\xi(\infty) = \infty$, a fairly direct numerical procedure will yield the minimizer of $\xi(z)$.

3. Some examples.

Example 1. For this example, take $G = G^+$ as in Section 1, so the problem is to find the best equivariant estimator of the lower triangular square root of Σ, say θ. The loss function in this example is taken to be related to that used by Stein [see James and Stein (1960)] in the estimation of Σ. The function K which gives the equivariant loss function L [via (2.3)] is

$$K_1(u) = \operatorname{tr} uu' - \log\det(uu') - p. \tag{3.1}$$

The function K_1 can be thought of as a measure of the distance from $u \in G^+$ to $e \in G^+$ because K_1 is uniquely minimized at $u = e$. The best equivariant estimator for this K_1 is obtained by minimizing

$$\begin{aligned}\mathscr{E}_e K_1(Xa) &= \mathscr{E}_e[\operatorname{tr}(Xaa'X') - \log\det(Xaa'X') - p]\\ &= \mathscr{E}_e[\operatorname{tr} X'Xaa' - \log\det aa' - \log\det(X'X) - p]\\ &= \operatorname{tr}(\mathscr{E}_e X'X)aa' - \log\det aa' + c,\end{aligned} \tag{3.2}$$

where c is a fixed constant. That $\log\det(X'X)$ has a finite expectation follows from (1.10) and the finiteness of $\mathscr{E}\log Z$, where Z is χ_m^2 and m is a positive integer.

From Lemma 2.1, the minimum is achieved at $B = aa' = (\mathscr{E}_e X'X)^{-1}$. Since

$$\mathscr{E}_e X'X = D_1, \tag{3.3}$$

where D_1 is diagonal with diagonal elements $d_i = n - 2i + p + 1$, $i = 1, \dots, p$ [when the original data are $\mathscr{W}(\Sigma, p, n)$], the minimizer is

$$a_0 = D_1^{-1/2} \in G^+.$$

Thus, the best equivariant estimator for this example is

$$t_1(X) = XD_1^{-1/2}. \tag{3.4}$$

EXAMPLE 2. This example is similar to that of Example 1 except that we take the function K [in (2.3)] to be

$$K_2(u) = \operatorname{tr}(uu')^{-1} - \log\det(uu')^{-1} - p.$$

Again, K_2 can be interpreted as measuring the distance of u from $e \in G^+$. The best equivariant estimator is found by minimizing (over $a \in G^+$)

$$\begin{aligned}\mathscr{E}_e K_2(Xa) &= \mathscr{E}_e\left[\operatorname{tr}(Xaa'X')^{-1} - \log\det(Xaa'X')^{-1} - p\right]\\ &= \operatorname{tr}\mathscr{E}_e(X'X)^{-1}(aa')^{-1} - \log\det(aa')^{-1} + c,\end{aligned} \tag{3.5}$$

where c is a constant. From Lemma 2.1, the minimum is achieved at $B = (aa')^{-1} = [\mathscr{E}_e(X'X)^{-1}]^{-1}$. But, as shown in the Appendix,

$$\mathscr{E}_e(X'X)^{-1} = D_2 \in G^+$$

is a diagonal matrix which can be calculated. Thus, the minimizer is

$$a_0 = D_2^{1/2}$$

and

$$t_2(X) = XD_2^{1/2} \tag{3.6}$$

is the best equivariant estimator for the loss determined by K_2.

EXAMPLE 3. Again we consider the situation of Examples 1 and 2 with K given by

$$K_3(u) = \operatorname{tr} uu' + \operatorname{tr}(uu')^{-1}. \tag{3.7}$$

Note that $K_3(u) = K_3(u^{-1})$, so that estimation of θ and θ^{-1} is considered simultaneously. Consequently, up to an additive constant,

$$K_3(u) = K_1(u) + K_2(u).$$

For this case, the best equivariant estimator is found by minimizing

$$\begin{aligned} \mathscr{E}_e K_3(Xa) &= \mathscr{E}_e\left[\operatorname{tr} Xaa'X' + \operatorname{tr}(Xaa'X')^{-1}\right] \\ &= \operatorname{tr}(\mathscr{E}_e X'X)aa' + \operatorname{tr}\mathscr{E}_e(X'X)^{-1}(aa')^{-1} \\ &= \operatorname{tr} D_1 aa' + \operatorname{tr} D_2(aa')^{-1}. \end{aligned} \tag{3.8}$$

From Lemma 2.2, the minimizer of (3.8) is

$$B = aa' = D_2^{1/2}\left(D_2^{1/2}D_1D_2^{1/2}\right)^{-1/2}D_2^{1/2} = D_2^{1/2}D_1^{-1/2},$$

so that

$$a_0 = \left(D_2D_1^{-1}\right)^{1/4},$$

and the best equivariant estimator is

$$t_3(X) = X\left(D_2D_1^{-1}\right)^{1/4}. \tag{3.9}$$

4. Invariant quadratic loss functions. In this section, we derive other best equivariant estimators, but for loss functions that are equivariant and quadratic [such as those discussed in Olkin and Selliah (1977)]. Before deriving these estimators, we require some preliminaries taken from Eaton (1970).

Consider a random vector Y taking values in an inner product space $(V, [\cdot, \cdot])$. Also, let

$$C(Y): V \to V \tag{4.1}$$

be an a.e. positive definite linear transformation [possibly random since $C(Y)$ can depend on Y]. For $a \in V$, the minimizer of

$$\psi(a) = \mathscr{E}[(a - Y), C(Y)(a - Y)] \tag{4.2}$$

is

$$a_0 = (\mathscr{E}C(Y))^{-1}\mathscr{E}(C(Y)Y). \tag{4.3}$$

The proof that a_0 is the minimizer of (4.2) consists in writing $\psi(a)$ as

$$\psi(a) = \psi(a_0) + \mathscr{E}[a - a_0, C(Y)(a - a_0)] + 2\mathscr{E}[a - a_0, C(Y)(a_0 - Y)]$$

and noting that the cross-product term vanishes. We have found that this vector space version of minimizing a quadratic function provides a useful way of looking at the next two examples as well as other problems in multivariate analysis. However, Lemma 2.3 also provides the minimizers in the following two examples.

EXAMPLE 4. Again take $X \in G^+$ as in Example 1. The function K which gives the equivariant loss function [via (2.3)] is taken to be

$$K_4(u) = \operatorname{tr}(u - I_p)(u - I_p)'. \tag{4.4}$$

To find the best equivariant estimator, the function

$$\psi_4(a) = \mathscr{E}K_4(Xa) = \mathscr{E}\,\mathrm{tr}(Xa - I_p)(Xa - I_p)' \tag{4.5}$$

needs to be minimized over $a \in G^+$. Here and in what follows $\mathscr{E}_e \doteq \mathscr{E}$.

Let V be the vector space of all $p \times p$ real matrices with the inner product

$$[v_1, v_2] = \mathrm{tr}\, v_1 v_2'.$$

Define $Y \in V$ by $Y = X^{-1}$, so $Y \in G^+ \subseteq V$, and define $C(Y)$ by

$$C(Y) = (YY')^{-1} \otimes I_p,$$

so that $C(Y)$ is a positive definite linear transformation on V to V. [For a discussion of the Kronecker product notation $\otimes$, see Eaton (1983), pages 34–36.] The value of $C(Y)$ at $v \in V$ is

$$C(Y)v = \left((YY')^{-1} \otimes I_p\right)v = (YY')^{-1}v.$$

In terms of Y, we have

$$\begin{aligned}\psi_4(a) &= \mathscr{E}\,\mathrm{tr}\, X(a - X^{-1})(a - X^{-1})'X' = \mathscr{E}\,\mathrm{tr}\, X'X(a - X^{-1})(a - X^{-1})' \\ &= \mathscr{E}[C(Y)(a - Y), (a - Y)] = \mathscr{E}[(a - Y), C(Y)(a - Y)],\end{aligned}$$

which is minimized at

$$\begin{aligned}a_0 &= (\mathscr{E}C(Y))^{-1}\mathscr{E}(C(Y)Y) \\ &= \left(\mathscr{E}\left[(YY')^{-1} \otimes I_p\right]\right)^{-1}\mathscr{E}\left[(YY')^{-1} \otimes I_p\right]Y \\ &= \left[(\mathscr{E}X'X) \otimes I_p\right]^{-1}\mathscr{E}X' = (\mathscr{E}X'X)^{-1}\mathscr{E}X'.\end{aligned}$$

Lemma 2.3 also shows a_0 is the minimizer of ψ_4. Now, $\mathscr{E}X' = \mathscr{E}X$ is known from (1.11) and $\mathscr{E}X'X = D_1$ was computed in (3.3) in Example 1, so that

$$a_0 = D_1^{-1}D,$$

and the best equivariant estimator is

$$t_4(X) = XD_1^{-1}D. \tag{4.6}$$

Example 5. With X and G^+ as in Example 4, consider a loss function determined by

$$K_5(u) = \mathrm{tr}\left(u^{-1} - I_p\right)\left(u^{-1} - I_p\right)'. \tag{4.7}$$

The best equivariant estimator is found by minimizing

$$\begin{aligned}\psi_5(a) &= \mathscr{E}K_5(Xa) = \mathscr{E}\,\mathrm{tr}(a^{-1}X^{-1} - I)(a^{-1}X^{-1} - I)' \\ &= \mathscr{E}\,\mathrm{tr}(a^{-1} - X)X^{-1}X^{-1\prime}(a^{-1} - X)'.\end{aligned} \tag{4.8}$$

With the vector space and inner product as in Example 4, take $Y = X$ and take

$$C(Y) = I \otimes (Y^{-1}Y^{-1\prime}).$$

Hence,

$$C(Y)v = v(Y^{-1}Y^{-1'})', \qquad v \in V,$$

is positive definite on V and

$$\psi_5(a) = \mathscr{E}\left[(a^{-1} - Y), C(Y)(a^{-1} - Y)\right].$$

The minimizing a^{-1} is

$$\begin{aligned}(a^{-1})_0 &= \left[\mathscr{E}\left(I_p \otimes (X'X)^{-1}\right)\right]^{-1}\mathscr{E}\left[I_p \otimes X^{-1}X^{-1'}\right]X \\ &= \left[I_p \otimes \left(\mathscr{E}(X'X)^{-1}\right)^{-1}\right]\mathscr{E}(X^{-1})' \\ &= \left[\mathscr{E}(X^{-1})\right]\left[\mathscr{E}(X'X)^{-1}\right]^{-1},\end{aligned}$$

or

$$a_0 = \left[\mathscr{E}(X'X)^{-1}\right]\left[\mathscr{E}(X^{-1})'\right]^{-1}.$$

Alternatively, Lemma 2.3 shows that a_0 minimizes ψ_5. The computation $\mathscr{E}(X'X)^{-1} = D_2$ was given in Example 2. Using a partitioning argument as in the first line of the proof of Lemma 2.3, it follows that $\mathscr{E}X^{-1}$ is diagonal with diagonal elements $\mathscr{E}x_{ii}^{-1}$, $i = 1, \dots, p$. Thus, the best equivariant estimator is

$$t_5(X) = XD_2D_5,$$

where

$$D_5 = \left[\mathscr{E}X^{-1}\right]^{-1} \tag{4.9}$$

is diagonal with diagonal elements

$$\left(\mathscr{E}x_{ii}^{-1}\right)^{-1} = \sqrt{2}\,\Gamma\left(\frac{n-i+1}{2}\right)\Big/\Gamma\left(\frac{n-i}{2}\right), \qquad i = 1, \dots, p.$$

EXAMPLE 6. With X and G^+ as in Examples 4 and 5, consider the loss function

$$K_6(u) = \operatorname{tr}(u - I_p)(u - I_p)' + \operatorname{tr}\left(u^{-1} - I_p\right)\left(u^{-1} - I_p\right)' = K_4(u) + K_5(u)$$

as a parallel to the loss function of Example 3. The best equivariant estimator is found by minimizing

$$\begin{aligned}\psi_6(a) &= \mathscr{E}K_6(Xa) \\ &= \mathscr{E}\operatorname{tr}(Xa - I_p)(Xa - I_p)' + \mathscr{E}\operatorname{tr}\left(a^{-1}X^{-1} - I_p\right)\left(a^{-1}X^{-1} - I_p\right)' \\ &= \operatorname{tr}(aa')(\mathscr{E}X'X) + \operatorname{tr}(aa')^{-1}\mathscr{E}(X'X)^{-1} \\ &\quad -2\operatorname{tr}a^{-1}(\mathscr{E}X^{-1}) - 2\operatorname{tr}a(\mathscr{E}X) + 2p \\ &\equiv \operatorname{tr}(aa')^{-1}D_2 + \operatorname{tr}(aa')D_1 - 2\operatorname{tr}aD_3 - 2\operatorname{tr}a^{-1}D_4 + 2p,\end{aligned}$$

where $D_1 = \mathscr{E}X'X$, $D_2 = \mathscr{E}(X'X)^{-1}$, $D_3 = \mathscr{E}X$, $D_4 = \mathscr{E}X^{-1}$ and each D_i is diagonal.

The minimizer a_0 of $\psi_6(a)$ is diagonal and is obtained as a consequence of Lemma 2.4. Thus, the best equivariant estimator is

$$t_6 = X\tilde{D},$$

where the elements of $\tilde{D}$ are obtained from Lemma 2.4.

5. Summary. For each of the loss functions given in Examples 1–6, we have obtained the best equivariant estimator to be of the form $t(x) = XD^*$, where D^* is diagonal. We now summarize these results in terms of the elements of the diagonal matrix D^*.

The unbiased estimator is given by $D^* = D^{-1}$, where the elements of D are defined in (1.12). The result of Example 1 yields $D^* = D_1^{-1/2}$, where the elements of D_1 are given in (3.3). The result of Example 2 yields $D^* = D_2^{1/2}$, where the elements of D_2 are given by (A.7). The results of Examples 3 and 4 are given by $D^* = D_2^{1/4}D_1^{-1/4}$ and $D_1^{-1}D$, respectively. The result of Example 5 is given by $D^* = D_2D_5$, where D_5 is obtained in (4.9). The result of Example 6 is a function of these various other diagonal matrices and must be obtained numerically from (2.8).

The improvement in risk obtained by using the best equivariant estimator rather than the unbiased estimator is

$$\Delta = \mathscr{E}_e K(XD^{-1}) - \mathscr{E}_e K(Xa_0),$$

where K specifies the loss function via (2.3) and a_0 is the minimizer of (2.7). The explicit computation of Δ is troublesome. For example, when K is given by K_1 in (3.1), then a routine computation using results herein yields

$$\Delta = \Sigma_1^p(v_i - 1 - \log v_i),$$

where

$$v_i = \frac{(n-i+1)\Gamma^2((n-i+1)/2)}{2\Gamma^2((n-i+2)/2)}, \qquad i = 1, \dots, p.$$

The numerical evaluation of Δ is certainly possible, but useful algebraic expressions for Δ have not been found.

APPENDIX

Let S have a $W(I_p, p, n)$ distribution and write

$$S = XX',$$

where X is lower triangular with positive diagonal elements. We want to compute the matrix

$$H = \mathscr{E}(X'X)^{-1}. \tag{A.1}$$

For any diagonal matrix D with ± 1 on the diagonal, DXD and X have the same distribution, so that

$$H = \mathscr{E}(X'X)^{-1} = \mathscr{E}[(DXD)'(DXD)]^{-1} = D\mathscr{E}(X'X)^{-1}D = DHD.$$

Hence, H must be a diagonal matrix, and only the expectations of the diagonal elements of $(X'X)^{-1}$ need to be calculated. We know that

$$\mathscr{L}(x_{ii}^2) = \chi_{n-i+1}^2,$$

where x_{ii} is the ith diagonal element of X. Hence,

$$\text{(A.2)} \qquad \alpha_i \equiv \mathscr{E}(x_{ii}^2)^{-1} = \frac{1}{n-i-1}, \qquad i = 1, \dots, p.$$

With X partitioned into blocks, we have

$$X^{-1} = \begin{pmatrix} X_{11} & 0 \\ X_{21} & X_{22} \end{pmatrix}^{-1} = \begin{pmatrix} X_{11}^{-1} & 0 \\ -X_{22}^{-1}X_{21}X_{11}^{-1} & X_{22}^{-1} \end{pmatrix}$$

and

$$(X'X)^{-1} = X^{-1}(X')^{-1} = \begin{pmatrix} X_{11}^{-1}(X_{11}')^{-1} & C \\ C' & A_{21}A_{21}' + X_{22}^{-1}(X_{22}')^{-1} \end{pmatrix},$$

where $A_{21} = -X_{22}^{-1}X_{21}X_{11}^{-1}$ and the matrix C does not enter into the computation. Taking X_{11} to be $(p-1) \times (p-1)$ and X_{22} a scalar,

$$H = \begin{pmatrix} \mathscr{E}(X'_{11}X_{11})^{-1} & 0 \\ 0 & \mathscr{E}(1/X_{22}^2)\left[1 + X_{21}(X'_{11}X_{11})^{-1}X_{21}'\right] \end{pmatrix}.$$

Let $h_1, \dots, h_p$ denote the diagonal elements of H. From the independence of the elements x_{ij} of X,

$$\text{(A.3)} \qquad h_p = \alpha_p\left[1 + \mathscr{E}X_{21}(X_{11}'X_{11})^{-1}X_{21}'\right],$$

with α_p given by (A.2). But

$$\begin{aligned} \mathscr{E}X_{21}(X_{11}'X_{11})^{-1}X_{21}' &= \mathscr{E}\operatorname{tr} X_{21}'X_{21}(X_{11}'X_{11})^{-1} \\ &= \operatorname{tr}(\mathscr{E}X_{21}'X_{21})\mathscr{E}(X_{11}'X_{11})^{-1} \end{aligned}$$

Since the elements of X_{21} are independent standard normal variates, $\mathscr{E}X_{21}'X_{21} = I_{p-1}$. Further,

$$\mathscr{E}(X_{11}'X_{11})^{-1} = H_{11},$$

where H_{11} is the upper left hand $(p-1) \times (p-1)$ corner of H. Thus, (A.3) becomes

$$\text{(A.4)} \qquad h_p = \alpha_p[1 + \operatorname{tr} H_{11}] = \alpha_p\left[1 + \Sigma_1^{p-1}h_i\right],$$

which together with $h_1 = \alpha_1$ yields the inductive equation

$$\text{(A.5)} \qquad h_j = \alpha_j\left[1 + \Sigma_1^{j-1}h_i\right], \qquad j = 1, \dots, p.$$

Solving (A.5) inductively yields

$$\text{(A.6)} \qquad \begin{aligned} h_1 &= \alpha_1, \\ h_j &= \alpha_j \prod_{i=1}^{j-1}(1 + \alpha_i), \qquad j = 2, \dots, p. \end{aligned}$$

Finally,

$$h_i = \frac{1}{n-i-1}\,\frac{n-1}{n-i}, \qquad i = 1, \ldots, p. \tag{A.7}$$

REFERENCES

BARTLETT, M. S. (1933). On the theory of statistical regression. *Proc. Roy. Soc. Edinburgh* **53** 260–283.

EATON, M. L. (1970). Some problems in covariance estimation (preliminary report). Technical Report 49, Dept. Statistics, Stanford Univ.

EATON, M. L. (1983). *Multivariate Statistics: A Vector Space Approach*. Wiley, New York.

JAMES, W. and STEIN, C. (1960). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 361–379. Univ. California Press.

KIEFER, J. (1957). Invariance, minimax sequential estimation, and continuous time processes. *Ann. Math. Statist.* **28** 573–601.

LEHMANN, E. (1983). *Theory of Point Estimation*. Wiley, New York.

MAHALANOBIS, P. C., BOSE, R. C and ROY, S. N. (1937). Normalisation of statistical variates and the use of rectangular co-ordinates in the theory of sampling distributions. *Sankhyā* **3** 1–40.

MARSHALL, A. W. and OLKIN, I. (1979). *Theory of Majorization and Its Applications*. Academic, New York.

OLKIN, I. (1985). Estimating the Cholesky decomposition. *Linear Algebra Appl.* **67** 201–205.

OLKIN, I. and SELLIAH, J. B. (1977). Estimating covariances in a multivariate normal distribution. In *Statistical Decision Theory and Related Topics, II* (S. S. Gupta and D. S. Moore, eds.) 313–326. Academic, New York.

DEPARTMENT OF STATISTICS
270 VINCENT HALL
UNIVERSITY OF MINNESOTA
206 CHURCH STREET, S. E.
MINNEAPOLIS, MINNESOTA 55455

DEPARTMENT OF STATISTICS
SEQUOIA HALL
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305