The Distance between Two Random Vectors with Given Dispersion Matrices

I. Olkin

Statistics Department
Sequoia Hall
Stanford University
Stanford, California 94305
and
Universität Freiburg im Breisgau

and

F. Pukelsheim

Institut für Mathematische Stochastik
Albert-Ludwigs-Universität
D-7800 Freiburg im Breisgau
Federal Republic of Germany

Submitted by Richard W. Cottle

ABSTRACT

For two p-dimensional random vectors X and Y with dispersion matrices Σ_{11} and Σ_{22}, respectively, we determine that covariance matrix Ψ_0 of X and Y that minimizes the L_2-distance between X and Y. There is a dual to this problem that is of interest in another context.

1. INTRODUCTION

Consider two p-variate normal distributions with zero means and positive definite dispersion matrices Σ_{11} and Σ_{22}. In the theory of strong approximations it is of interest to construct p-dimensional random vectors X and Y distributed according to $N(0, \Sigma_{11})$ and $N(0, \Sigma_{22})$, respectively, such that the L_2-distance between X and Y is minimal.

Let

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Psi \\ \Psi' & \Sigma_{22} \end{pmatrix}$$

52 Vanderbilt Ave., New York, NY 10017 0024-3795/82/070257-07$2.75

denote the dispersion matrix of (X,Y). Then the problem is to minimize $E\,\mathrm{tr}(X-Y)'(X-Y)=\mathrm{tr}(\Sigma_{11}+\Sigma_{22}-2\Psi)$. The restriction that Σ be nonnegative definite is equivalent to requiring that the Schur complement $\Sigma_{11}-\Psi\Sigma_{22}^{-1}\Psi'$ be nonnegative definite. Consequently, under the assumption that $\Sigma_{11}>0$, $\Sigma_{22}>0$, the extremal problem becomes

$$\max_{\Sigma_{11}-\Psi\Sigma_{22}^{-1}\Psi'\geqslant 0} \mathrm{tr}\,2\Psi, \tag{1a}$$

where the Loewner ordering $A \geqslant B$ ($A>B$) means that $A-B$ is nonnegative (positive) definite.

As a consequence of Theorem 3 we show that the problem (1) has a unique solution

$$\Psi_0=\Sigma_{11}\Sigma_{22}^{1/2}\left(\Sigma_{22}^{1/2}\Sigma_{11}\Sigma_{22}^{1/2}\right)^{-1/2}\Sigma_{22}^{1/2}, \tag{2}$$

where $A^{1/2}$ denotes the unique positive definite square root of the positive definite matrix A.

When $p=1$, $\Psi_0=\sqrt{\sigma_{11}\sigma_{22}}$, so that the correlation between X and Y is 1. When Σ_{11} and Σ_{22} are diagonal, Ψ_0 is diagonal, so that $(X_1,Y_1),\ldots,(X_p,Y_p)$ are independent bivariate random vectors that are perfectly correlated.

When $X_1,\ldots,X_p$ and $Y_1,\ldots,Y_p$ are equicorrelated with correlations ρ and η, respectively, then Ψ_0 is a matrix with equal diagonal elements $\psi_{ii}=a(\rho,\eta)$, $i=1,\ldots,p$, and equal off-diagonal elements $\psi_{ij}=b(\rho,\eta)$, $i\neq j=1,\ldots,p$. The constants $a(\rho,\eta)$ and $b(\rho,\eta)$ are rather complicated functions of ρ and η.

We actually obtain a stronger result by assuming only nonnegative definiteness of Σ_{11} and Σ_{22}. This permits a comparison of random vectors X and Y of different lengths by appropriately including random variables degenerate at zero.

2. A DUALITY THEOREM

In another context Anderson and Olkin [1] consider the extremal problem

$$\min_{S>0} \mathrm{tr}\left(\Sigma_{11}S+\Sigma_{22}S^{-1}\right), \qquad \Sigma_{11}>0, \quad \Sigma_{22}>0. \tag{1b}$$

It is interesting to note that the problems (1a) and (1b) are dual to each other. As a consequence we have a particularly simple way to investigate the optimal solutions. Since the set of matrices satisfying $\Sigma_{11}-\Psi\Sigma_{22}^{-1}\Psi'\geqslant 0$ forms a convex set (see [2, p. 468]), the problem can also be considered from the more general programming theory as outlined in [3]. However, we need not use this route, since a direct argument establishes the duality.

Our first theorem shows the interplay of the problems (1a) and (1b). However, we first require the following lemma.

LEMMA 1. *Let $\Sigma_{11}\geqslant 0$, $\Sigma_{22}\geqslant 0$. Then*

$$\Sigma\equiv\begin{pmatrix}\Sigma_{11} & \Psi\\ \Psi' & \Sigma_{22}\end{pmatrix}\geqslant 0$$

if and only if $\Psi\in\Omega$, where

$$\Omega=\{\Psi:\Sigma_{22}\Sigma_{22}^{-}\Psi'=\Psi',\ \Sigma_{11}-\Psi\Sigma_{22}^{-}\Psi'\geqslant 0\}, \tag{3}$$

and Σ_{22}^{-} is any generalized inverse of Σ_{22}.

Proof. For the direct part, assume $\Sigma\geqslant 0$ and suppose the vector y is in the nullspace of Σ_{22}. Then letting $\lambda\neq 0$ tend to zero in

$$0\leqslant(\lambda x',\lambda^{-1}y)\Sigma(\lambda x,\lambda^{-1}y)'=\lambda^2x'\Sigma_{11}x+2x'\Psi y$$

proves that $\Psi y=0$. Thus, $\mathrm{nullspace}(\Sigma_{22})\subseteq\mathrm{nullspace}(\Psi)$, or equivalently, $\mathrm{range}(\Psi')\subseteq\mathrm{range}(\Sigma_{22})$, which in turn is equivalent to $\Sigma_{22}\Sigma_{22}^{-}\Psi'=\Psi'$.

The second property is implied by

$$0\leqslant\begin{pmatrix}I & -\Psi\Sigma_{22}^{-}\\ 0 & I\end{pmatrix}\begin{pmatrix}\Sigma_{11} & \Psi\\ \Psi' & \Sigma_{22}\end{pmatrix}\begin{pmatrix}I & 0\\ -\Sigma_{22}^{-}\Psi' & I\end{pmatrix}=\begin{pmatrix}\Sigma_{11}-\Psi\Sigma_{22}^{-}\Psi' & 0\\ 0 & \Sigma_{22}\end{pmatrix}.$$

The converse part follows from

$$0\leqslant\begin{pmatrix}I & \Psi\Sigma_{22}^{-}\\ 0 & I\end{pmatrix}\begin{pmatrix}\Sigma_{11}-\Psi\Sigma_{22}^{-}\Psi' & 0\\ 0 & \Sigma_{22}\end{pmatrix}\begin{pmatrix}I & 0\\ \Sigma_{22}^{-}\Psi' & I\end{pmatrix}=\Sigma. \qquad \blacksquare$$

As a consequence of Lemma 1, under the weaker assumption $\Sigma_{11}^{-}\geqslant 0$, $\Sigma_{22}\geqslant 0$, the problem (1a) generalizes to

$$\max_{\Psi\in\Omega}\mathrm{tr}\,2\Psi. \tag{1a'}$$

THEOREM 2 (Matrix inequality). *For $\Psi\in\Omega$ and for an arbitrary $p\times p$ matrix R with generalized inverse R^{-},*

$$R'\Sigma_{11}R+R^{-}\Sigma_{22}(R^{-})'\geqslant R'\Psi(R^{-})+R^{-}\Psi'R, \tag{4}$$

with equality if and only if

$$R'(\Sigma_{11} - \Psi\Sigma_{22}^{-}\Psi')R = 0, \tag{5a}$$

$$R'\Psi = R^{-}\Sigma_{22}. \tag{5b}$$

Proof. Let G denote any generalized inverse of R. The inequality (4) follows from the nonnegative definiteness of

$$(R', -G)\begin{pmatrix} \Sigma_{11} & \Psi \\ \Psi' & \Sigma_{22} \end{pmatrix}\begin{pmatrix} R \\ -G' \end{pmatrix}.$$

To show the case of equality, let

$$T = R'\Psi(\Sigma_{22}^{1/2})^{-} - G\Sigma_{22}^{1/2}$$

Then equality in (4) holds if and only if equality holds in

$$\begin{aligned} R'\Sigma_{11}R + G\Sigma_{22}G' &\geqslant R'\Sigma_{11}R + G\Sigma_{22}G' - TT' \\ &= R'\Psi G' + G\Psi' R + R'(\Sigma_{11} - \Psi\Sigma_{22}^{-}\Psi')R \\ &\geqslant R'\Psi G' + G\Psi' R. \end{aligned}$$

In the case of equality, (5a) holds and $T = 0$, which is equivalent to (5b). ∎

In order to connect Theorem 2 with the problem (1b), let $S = RR'$ satisfy the condition $\mathrm{range}(\Sigma_{22}) \subseteq \mathrm{range}(S)$, that is,

$$SS^{-}\Sigma_{22} = \Sigma_{22},$$

where S^{-} is any generalized inverse of S. Accordingly, assuming merely $\Sigma_{11} \geqslant 0$, $\Sigma_{22} \geqslant 0$, we modify the problem (1b) to

$$\min_{s \in \mathbb{S}} \mathrm{tr}(\Sigma_{11}S + \Sigma_{22}S^{-}), \tag{1b$'$}$$

where

$$\mathbb{S} = \{S : S \geqslant 0,\ SS^{-}\Sigma_{22} = \Sigma_{22}\}. \tag{6}$$

Corollary 3 (Mutual boundedness). *For all matrices $\Psi \in \Omega$ and $S \in \mathbb{S}$, the inequality*

$$\mathrm{tr}(\Sigma_{11}S + \Sigma_{22}S^{-}) \geqslant \mathrm{tr}\,2\Psi \tag{7}$$

holds, with equality if and only if

$$S\Sigma_{11}S = \Sigma_{22} = S\Psi. \tag{8}$$

Proof. Fix S^{-} and $S = RR'$. Then $G = R'S^{-}$ is a generalized inverse of R, and by (6),

$$\Sigma_{22}G'G = \Sigma_{22}S^{-}RR'S^{-} = \Sigma_{22}S^{-}SS^{-} = \Sigma_{22}S^{-}.$$

Hence, from (4)

$$\begin{aligned} \mathrm{tr}(\Sigma_{11}S + \Sigma_{22}S^{-}) &= \mathrm{tr}(\Sigma_{11}RR' + \Sigma_{22}G'G) \\ &\geqslant \mathrm{tr}(\Psi G'R' + RG\Psi') = 2\,\mathrm{tr}\,RG\Psi' = 2\,\mathrm{tr}\,SS^{-}\Psi'. \end{aligned}$$

But, because $\Psi \in \Omega$ and (6) holds,

$$\Psi' = \Sigma_{22}\Sigma_{22}^{-}\Psi' = SS^{-}\Sigma_{22}\Sigma_{22}^{-}\Psi' = SS^{-}\Psi'. \tag{9}$$

The condition (5b) implies that $S\Psi = \Sigma_{22}$, which when inserted in (5a) yields $S\Sigma_{11}S = \Sigma_{22}$. Thus equality in (7) forces (8), whereas the converse is immediate from (9). ∎

From (8) it follows that equality in (7) holds only if $\mathrm{rank}(\Sigma_{22}) = \mathrm{rank}(S\Sigma_{11}S) \leqslant \mathrm{rank}(\Sigma_{11})$. The rank assumption can be made without loss of generality; however, we require the slightly stronger assumption that $\mathrm{range}(\Sigma_{22}) \subseteq \mathrm{range}(\Sigma_{11})$ in the following duality theorem.

Theorem 4 (Duality). *If $\mathrm{range}(\Sigma_{22}) \subseteq \mathrm{range}(\Sigma_{11})$ then the problems (1a$'$) and (1b$'$) share the same optimal value*

$$\max_{\Psi \in \Omega} \mathrm{tr}\,2\Psi = \min_{S \in \mathbb{S}} \mathrm{tr}(\Sigma_{11}S + \Sigma_{22}S^{-}) = 2\,\mathrm{tr}\left(\Sigma_{22}^{1/2}\Sigma_{11}\Sigma_{22}^{1/2}\right)^{1/2}, \tag{10}$$

with solutions

$$\Psi_0 = \Sigma_{11}\Sigma_{22}^{1/2}\left[\left(\Sigma_{22}^{1/2}\Sigma_{11}\Sigma_{22}^{1/2}\right)^{1/2}\right]^{-}\Sigma_{22}^{1/2} \equiv \Sigma_{11}S_0, \tag{11a}$$

$$S_0 = \Sigma_{22}^{1/2}\left[\left(\Sigma_{22}^{1/2}\Sigma_{11}\Sigma_{22}^{1/2}\right)^{1/2}\right]^{-}\Sigma_{22}^{1/2}. \tag{11b}$$

If $\mathrm{range}(\Sigma_{22}) = \mathrm{range}(\Sigma_{11})$, then Ψ_0 is the unique solution of the problem (1a$'$), and S_0 is the unique solution of problem (1b$'$) that satisfies $\mathrm{range}(S_0) = \mathrm{range}(\Sigma_{22})$.

Proof. Because of the assumption $\text{range}(\Sigma_{22}) \subseteq \text{range}(\Sigma_{11})$, the matrices Ψ_0 and S_0 are invariant under the choice of the generalized inverse $\left[(\Sigma_{22}^{1/2}\Sigma_{11}\Sigma_{22}^{1/2})^{1/2}\right]^-$. The matrices Ψ_0 and S_0 defined by (11) are clearly feasible for the problems (1a′) and (1b′) and satisfy (8), thus proving (10).

To establish uniqueness, an argument similar to that in the proof of Lemma 1 shows that any $\Psi \in \Omega$ satisfies $\text{range}(\Psi) \subseteq \text{range}(\Sigma_{11})$. Hence, if $\text{range}(\Sigma_{11}) = \text{range}(\Sigma_{22})$ and $S \in \mathcal{S}$, then $\text{range}(\Psi) \subseteq \text{range}(S)$, and $SS^-\Psi = \Psi$. In particular, from (8), any two optimal solutions Ψ_0 and $\tilde{\Psi}_0$ of problem (1′) satisfy $S_0\Psi_0 = \Sigma_{22} = S_0\tilde{\Psi}_0$, which yields

$$\Psi_0 = S_0S_0^-\Psi_0 = S_0^-S_0\Psi_0 = S_0^-\Sigma_{22} = S_0^-S_0\tilde{\Psi}_0 = S_0S_0^-\tilde{\Psi}_0 = \tilde{\Psi}_0.$$

Finally, any two optimal solutions S_0 and $\tilde{S}_0$ of (1b′) satisfy $S_0\Psi_0 = \Sigma_{22} = \tilde{S}_0\Psi_0$. Suppose, in addition, that $\text{range}(S_0) = \text{range}(\tilde{S}_0) = \text{range}(\Sigma_{22})$. From the above we have that $\text{range}(\Psi) \subseteq \text{range}(\Sigma_{22})$, whereas (8) implies that $\text{rank}(\Sigma_{22}) \leqslant \text{rank}(\Psi)$. Thus, $\text{range}(\Sigma_{22}) = \text{range}(\Psi_0)$, and

$$S_0 = S_0\Psi\Psi_0^- = \Sigma_{22}\Psi_0^- = \tilde{S}_0\Psi_0\Psi_0^- = \tilde{S}_0. \quad \blacksquare$$

NOTE. If $z \perp \text{range}(\Sigma_{22}) = \text{range}(\Sigma_{11})$, then $S_0 + zz'$ also satisfies (8), and hence is also optimal, which shows that the condition $\text{range}(S_0) = \text{range}(\Sigma_{22})$ cannot be relaxed.

REMARK. When $\text{range}(\Sigma_{11}) = \text{range}(\Sigma_{22})$, the problem (1a′) can also be formulated with the Schur complement $\Sigma_{22} - \Psi'\Sigma_{11}^-\Psi$. Accordingly, an alternative representation of its unique optimal solution is

$$\Psi_0' = \Sigma_{22}\Sigma_{11}^{1/2}\left[(\Sigma_{11}^{1/2}\Sigma_{22}\Sigma_{11}^{1/2})^{1/2}\right]^-\Sigma_{11}^{1/2}.$$

If $\lambda_1(A),\dots,\lambda_p(A)$ denote the characteristic roots of A, then the optimal value (9) becomes

$$2\,\text{tr}(\Sigma_{22}^{1/2}\Sigma_{11}\Sigma_{22}^{1/2})^{1/2} = 2\sum_1^p \lambda_i\left[(\Sigma_{22}^{1/2}\Sigma_{11}\Sigma_{22}^{1/2})^{1/2}\right]$$

$$= 2\sum_1^p \lambda_i^{1/2}(\Sigma_{22}^{1/2}\Sigma_{11}\Sigma_{22}^{1/2}) = 2\sum_1^p \lambda_i^{1/2}(\Sigma_{11}\Sigma_{22}),$$

which provides a more symmetric expression in Σ_{11} and Σ_{22}.

Note added in proof. During the proof stage the authors note the appearance of a paper dealing with the same problem, though motivated from a slightly different point of view. The reference is D. C. Dowson and B. V. Landau, the Fréchet distance between multivariate normal distributions, J. **Multivariate Anal.** 12:450–455 (1982).

We are grateful to E. Berger for calling this problem to our attention, and to Kai-Tai Fang for his comments and suggestions.

REFERENCES

1 T. W. Anderson and I. Olkin, An extremal problem for positive definite matrices, *Linear and Multilinear Algebra* 6:257–262 (1978).

2 A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*, Academic, New York, 1979.

3 F. Pukelsheim, A quick introduction to mathematical programming with applications to most powerful tests, nonnegative variance estimation, and optimal design theory, Technical Report No. 128, Dept. of Statistics, Stanford Univ., 1978.

Received 8 March 1982; revised 6 July 1982