Nonnegative definiteness of the estimated dispersion matrix in a multivariate linear model

Friedrich Pukelsheim, George P. H. Styan

Angaben zur Veröffentlichung / Publication details:

Pukelsheim, Friedrich, and George P. H. Styan. 1979. "Nonnegative definiteness of the estimated dispersion matrix in a multivariate linear model." *Bulletin de l'Académie Polonaise des Sciences: Série des Sciences Mathématiques* 27 (3-4): 327–30.

BULLETIN DE L'ACADÉMIE
POLONAISE DES SCIENCES
Série des sciences mathématiques
Vol. XXVII, No. 3-4, 1979

STATISTICS

Nonnegative Definiteness of the Estimated Dispersion Matrix in a Multivariate Linear Model

by

Friedrich PUKELSHEIM and George P. H. STYAN

Presented by K. URBANIK on April 13, 1978

Summary. Estimation is considered in a model where both the mean vector and the dispersion matrix have linear decompositions. It is shown that after an invariance reduction with respect to mean translation, MINQUE provides a nonnegative definite estimate of the dispersion matrix, when the decomposing matrices span a quadratic subspace of symmetric matrices. With normality, MINQUE is seen to equal the restricted maximum likelihood estimate and to be of uniformly minimum variance.

1. Introduction. Consider independent and identically distributed random $\mathbb{R}^n$-vectors $\mathbf{Y}_\alpha$, $\alpha=1, \ldots, N$, with common mean vector $\sum_{\pi=1}^{p} b_\pi \mathbf{x}_\pi$ and common dispersion matrix $\sum_{\kappa=1}^{k} t_\kappa \mathbf{V}_\kappa$, where interest concentrates on estimating the vector $\mathbf{t}:=(t_1, \ldots, t_k)'$ of dispersion coefficients. Various procedures have been put forward and discussed in the literature: (i) minimum norm unbiased quadratic invariant estimation (MINQUE, C.R. Rao [8, p. 302]), and, under normality, (ii) uniform minimum variance unbiased invariant estimation (UMVU, Seely [9]), and (iii) restricted (by invariance) maximum likelihood estimation (REML, Corbeil, Searle [2]). In this paper invariance is to be understood with respect to the group of all mean translations $\{\mathbf{y}\to\mathbf{y}+\Sigma b_\pi \mathbf{x}_\pi \mid (b_1, \ldots, b_p)' \in \mathbb{R}^p\}$, a maximal invariant statistic being $\mathbf{MY}$ where $\mathbf{M}$ projects orthogonally onto the orthogonal complement of the space spanned by $\mathbf{x}_1, \ldots, \mathbf{x}_p$; hence reduction by invariance yields the residual vectors $\mathbf{MY}_\alpha$ with mean $\mathbf{0}$ and dispersion matrix $\Sigma t_\kappa \mathbf{MV}_\kappa \mathbf{M}$.

Our main result may be roughly summarized as follows: If estimates according to each of the three procedures above exist, then they coincide, and the common estimate $\hat{\mathbf{t}}$ yields a nonnegative definite estimate $\Sigma \hat{t}_\kappa \mathbf{MV}_\kappa \mathbf{M}$ of the dispersion matrix in the invariance reduced model. This holds true for any *finite* sample size $N \geqslant v:=\operatorname{rank} \mathbf{M}$, in contrast to asymptotic results on consistency as $N\to\infty$, cf. Anderson [1].

In Section 2, the invariance reduced model is discussed in a normal setting, and Section 3 is concerned with the linear model situation.

*) Address for correspondence: George P. H. Styan, Department of Mathematics, McGill University, Burnside Hall, 805 Sherbrooke Street West, Montréal, Québec, Canada H3A 2K6.

The vital assumption is the condition of Seely [9] that $\mathbf{MV}_1\mathbf{M}, \ldots, \mathbf{MV}_k\mathbf{M}$ span a k-dimensional quadratic subspace $\mathscr{B}$ of symmetric $n\times n$ matrices. The subspace $\mathscr{B}$ is quadratic if and only if $\mathbf{A}^2\in\mathscr{B}$ whenever $\mathbf{A}\in\mathscr{B}$, i.e., $\mathscr{B}$ is closed under the multiplication $\mathbf{A}\circ\mathbf{B}:=\frac{1}{2}(\mathbf{AB}+\mathbf{BA})$. Jensen [4] points out that the latter property makes $\mathscr{B}$ into a k-dimensional special Jordan algebra, and we shall adopt this more informative terminology. For a discussion with no initial invariance reduction see Gnot, Klonecki, Zmyślony [3].

2. The Normal Model. We will use the isomorphism *vec* that maps a matrix into a vector by ordering its entries lexicographically, see Pukelsheim [7].

THEOREM 1. *Consider independent and identically normally distributed random* R^{ν}*-vectors* $\mathbf{Z}_\alpha$ *with common mean* $\mathbf{0}$ *and common dispersion matrix* $\Sigma t_\kappa \mathbf{W}_\kappa$, *where* $N\geqslant\nu$. *Assume that the k decomposing matrices* $\mathbf{W}_\kappa$ *span a k-dimensional special Jordan algebra* $\mathscr{B}$. *Define* $G\subset\mathsf{R}^k$ *to be the region of those values* $\mathbf{t}$ *of the dispersion parameter such that* $\Sigma t_\kappa \mathbf{W}_\kappa$ *is positive definite, and assume* $G\neq\emptyset$. *Then*:

(a) *The maximum likelihood estimator for* $\mathbf{t}\in G$ *is almost surely equal to the uniform minimum variance unbiased estimator* $\hat{\mathbf{t}}:=(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\cdot\operatorname{vec}\mathbf{S}$, *where* $\mathbf{D}:=[\operatorname{vec}\mathbf{W}_1:\ldots:\operatorname{vec}\mathbf{W}_k]$, *and* $\mathbf{S}:=\Sigma\mathbf{Z}_\alpha\mathbf{Z}_\alpha'/N$.

(b) *The estimated dispersion matrix* $\hat{\mathbf{W}}:=\Sigma\hat{t}_\kappa\mathbf{W}_\kappa$ *is nonnegative definite; in fact, if the sample dispersion matrix* $\mathbf{S}$ *is positive definite, so is* $\hat{\mathbf{W}}$.

Proof. (a) Since G is open and connected it is a region, and its boundary ∂G consists of those $\mathbf{t}\in\mathsf{R}^k$ such that $\Sigma t_\kappa\mathbf{W}_\kappa$ is nonnegative definite and singular. The sample dispersion matrix $\mathbf{S}$ is almost surely positive definite. If $\mathbf{t}$ tends to ∂G, or $\|\mathbf{t}\|$ tends to ∞, the likelihood function L tends to zero [1, p. 5]. Since L is positive in G there exists a maximum in G, and no maximum lies on the boundary ∂G. Hence the maximum likelihood estimate is a solution of the likelihood equations

$$\mathbf{D}'\mathbf{F}^{-1}\mathbf{D}\hat{\mathbf{t}}=\mathbf{D}'\mathbf{F}^{-1}\operatorname{vec}\mathbf{S}, \tag{1}$$

where the matrix of fourth moments

$$\mathbf{F}=\mathbf{F}(\hat{\mathbf{t}}):=(\Sigma\hat{t}_\kappa\mathbf{W}_\kappa)\otimes(\Sigma\hat{t}_\kappa\mathbf{W}_\kappa). \tag{2}$$

If $\mathbf{F}$ in (1) is put equal to $\mathbf{F}(\mathbf{t}_0)$ for some given $\mathbf{t}_0\in G$, then (1) is a set of weighted normal equations, cf. [7, p. 628], and hence yields a minimum variance unbiased estimator for the vector parameter $\mathbf{t}$. Since the matrices $\mathbf{W}_\kappa$ span a special Jordan algebra, there exists an almost surely unique *uniform* minimum variance unbiased invariant estimator which does not depend on the choice of $\mathbf{t}_0\in G$. Thus

$$\hat{\mathbf{t}}=(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\operatorname{vec}\mathbf{S}, \tag{3}$$

since $G\neq\emptyset$ implies the existence of a nonsingular matrix $\mathbf{B}\in\mathscr{B}$, and so $\mathbf{B}^{-1}\in\mathscr{B}$ and $\mathbf{I}_\nu=\mathbf{B}\circ\mathbf{B}^{-1}\in\mathscr{B}$; the matrix $\mathbf{F}$ in (1) may, therefore, be set equal to $\mathbf{I}_{\nu^2}=\mathbf{I}_\nu\otimes\mathbf{I}_\nu$.

(b) As a linear operator on the space of symmetric matrices, $\hat{\mathbf{t}}$ is surjective and hence open, and so if for some positive definite matrix $\mathbf{S}_0$ the value $\hat{\mathbf{t}}(\mathbf{S}_0)\notin G$, the same is true for an open neighbourhood of S_0, i.e., for a set of positive Lebesgue measure. This contradicts part (a) that $\hat{\mathbf{t}}$ maps into G almost surely. For a singular sample dispersion matrix S, consider the limit $\mathbf{S}+\varepsilon\mathbf{I}_v$ as ε tends to zero. Q.E.D.

Part (a) may also be obtained from a reparametrization by $\boldsymbol{\theta}=\boldsymbol{\theta}(\mathbf{t})$, where the bijection $\boldsymbol{\theta}$ from G onto G solves $\Sigma\theta_\kappa(\mathbf{t})\mathbf{W}_\kappa=(\Sigma t_\kappa \mathbf{W}_\kappa)^{-1}$, as introduced by Seely [9, p. 715]. In this case one obtains an exponential family in the vector parameter $\boldsymbol{\theta}$ and standard theory applies, cf. Anderson [1]. A theorem proved by Mäkeläinen, Schmidt, Styan [6] may be used to obtain uniqueness of the solution to the likelihood equations (1).

3. The Multivariate Linear Model. We now return to the linear, but not necessarily normal, model discussed in Section 1.

THEOREM 2. *Consider independent and identically distributed random* $\mathbb{R}^n$*-vectors* $\mathbf{Y}_\alpha$, $\alpha=1, \ldots, N$, *with common mean vector* $\Sigma b_\pi \mathbf{x}_\pi$ *and common dispersion matrix* $\Sigma t_\kappa \mathbf{V}_\kappa$, *where* $N \geqslant v = \operatorname{rank} \mathbf{M}$. *Assume that the k matrices* $\mathbf{M}\mathbf{V}_\kappa\mathbf{M}$ *span a k-dimensional special Jordan algebra* $\mathcal{B}$ *that contains* $\mathbf{M}$. *Let* $\mathbf{D}_{\mathbf{M}} := [\operatorname{vec} \mathbf{M}\mathbf{V}_1\mathbf{M} : \ldots : \operatorname{vec} \mathbf{M}\mathbf{V}_k\mathbf{M}]$. *Then the* MINQUE

$$\hat{\mathbf{t}} = (\mathbf{D}_{\mathbf{M}}'\mathbf{D}_{\mathbf{M}})^{-1}\mathbf{D}_{\mathbf{M}}' \cdot \operatorname{vec} \mathbf{S} \tag{4}$$

for $\mathbf{t}$ *yields a nonnegative definite estimate* $\Sigma \hat{t}_\kappa \mathbf{M}\mathbf{V}_\kappa\mathbf{M}$ *of the invariance reduced dispersion matrix, this estimate being of rank* v *if* $\mathbf{S} := \Sigma\mathbf{M}\mathbf{Y}_\alpha(\mathbf{M}\mathbf{Y}_\alpha)'/N$ *is of rank* v.

Proof. It is easily checked that $\hat{\mathbf{t}}$ is the MINQUE in the enlarged model $[\mathbf{Y}_1'\mathbf{M} : \ldots : \mathbf{Y}_N'\mathbf{M}]' \sim (\mathbf{0}, \Sigma t_\kappa \mathbf{I}_N \otimes \mathbf{M}\mathbf{V}_\kappa\mathbf{M})$. The rest will be proved by reference to Theorem 1. Choose an $n \times v$ full rank v factor $\mathbf{Q}$ of $\mathbf{M}$, i.e., $\mathbf{M}=\mathbf{Q}\mathbf{Q}'$ and $\mathbf{Q}'\mathbf{Q}=\mathbf{I}_v$; then $\mathbf{Q}'\mathbf{Y}$ is another maximal invariant statistic [5, p. 707]. For the sole reason of proof, add a normality assumption. Then Theorem 1 is applicable to $\mathbf{Z}_\alpha := \mathbf{Q}'\mathbf{Y}_\alpha$, and yields the same $\hat{\mathbf{t}}$ as in (4); and the results on $\Sigma\hat{t}_\kappa \mathbf{Q}'\mathbf{V}_\kappa\mathbf{Q}$ imply the assertions on $\Sigma\hat{t}_\kappa \mathbf{M}\mathbf{V}_\kappa\mathbf{M}$. Q.E.D.

If a normality assumption is added to Theorem 2, then using Theorem 1, we obtain the following:

COROLLARY. *If the common distribution of* $\mathbf{Y}_1, \ldots, \mathbf{Y}_N$ *is normal, then* $\hat{\mathbf{t}}$ *is the* UMVU *and* REML *estimate of* $\mathbf{t}$, *as well as the* MINQUE.

Examples may be found in Corbeil and Searle [2]. In each one of their four cases a special Jordan algebra is present: equality of MINQUE (i.e., ANOVA estimators) and REML is implied by the Corollary and need not be checked explicitly, nor need the likelihood equations be solved iteratively.

This paper was presented at the Instytut Matematyczny PAN, Wrocław, by the first author, who would like to thank the Polish Academy of Sciences for their kind invitation. The authors would also like to thank T. W. Anderson, K. Conradsen, and H. Witting for very helpful discussions. The second author's research was begun while he was a Canada Council fellow at the University of Helsinki.

INSTITUT FÜR MATHEMATISCHE STOCHASTIK, ALBERT-LUDWIGS-UNIVERSITÄT, D-7800 FREIBURG IM BREISGAU (FEDERAL REPUBLIC OF GERMANY) AND DEPARTMENT OF STATISTICS, STANFORD UNIVERSITY, STANFORD, CALIFORNIA 94305 (USA)

DEPARTMENT OF MATHEMATICS, McGILL UNIVERSITY, MONTREAL, QUEBEC, H3A 2K6 (CANADA)

REFERENCES

[1] T. W. Anderson, *Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices*, pp. 1–24 in: R. C. Bose, I. M. Chakravarti, P. C. Mahalanobis, C. R. Rao, K. J. C. Smith (eds.), *Essays in Probability and Statistics*, Statistical Publishing Society, Calcutta, 1969, and The University of North Carolina Press, Chapel Hill, 1970.

[2] R. R. Corbeil, S. R. Searle, *A comparison of variance component estimators*, Biometrics, **32** (1976), 779–791.

[3] S. Gnot, W. Klonecki, R. Zmyślony, *Uniformly minimum variance unbiased estimation in Euclidean vector spaces*, Bull. Acad. Polon. Sci., Sér. Sci. Math. Astronom. Phys., **24** (1976), 281–286.

[4] S. T. Jensen, *Covariance hypotheses which are linear in both the covariance and the inverse covariance*, Preprint 1975, No. 1, Institute of Mathematical Statistics, University of Copenhagen, January, 1975.

[5] J. Kleffe, *A note on MINQUE for normal models*, Math. Operationsforsch. Statist., **7** (1976), 707–714.

[6] T. Mäkeläinen, K. Schmidt, G. P. H. Styan, *On the uniqueness of the maximum likelihood estimator of a vector valued parameter*, Discussion Paper No. 140, Indian Statistical Institute, Delhi, 1976.

[7] F. Pukelsheim, *Estimating variance components in linear models*, J. Multivariate Anal., **6** (1976), 626–629.

[8] C. R. Rao, *Linear statistical inference and its applications*, Second Edition, John Wiley, New York, 1973.

[9] J. Seely, *Quadratic subspaces and completeness*, Ann. Math. Statist., **42** (1971), 710–721.

Ф. Пукельшейм, Д. П. Х. Стыан, **Неотрицательная определенность оценивания дисперсионной матрицы в многомерной линейной модели**

Содержание. В работе рассматривается оценка в модели, где средний вектор и дисперсионная матрица обладают линейными разложениями. Показано, что после инвариантности редукции по отношению к среднему переносу, MINQUE даёт неотрицательную определенную оценку дисперсионной матрицы, когда разложимые матрицы охватывают квадратные подпространства симметричных матриц. По нормальности, MINQUE считается равным ограниченному наибольшему правдоподобию оценки и есть равномерным минимумом дисперсии.