Journal of Statistical Planning and Inference 12 (1985) 353-357
North-Holland

ON THE NONSINGULARITY OF PRINCIPAL SUBMATRICES OF A RANDOM ORTHOGONAL MATRIX

Takeaki KARIYA*

Institute of Economic Research, Hitotsubashi University, Kunitachi-shi, Tokyo 186, Japan.

C.F.J. WU**

Department of Statistics, University of Wisconsin, Madison, WI 53706, USA

Received 22 October 1984; revised manuscript received 19 April 1985
Recommended by M.L. Puri

Abstract: Let S be a random nonnegative definite matrix and let G be an orthogonal matrix such that $S=GDG'$, where D is the diagonal matrix of the latent roots of S. In this note, we prove the a.s. existence of G_{11}^{-1}, where $G=\begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}$, under some weak conditions on the distribution of S.

AMS Subject Classification: Primary 62H10, 60D05.

Key words and phrases: Nonsingularity of a submatrix of a random orthogonal matrix; Sample covariance matrix; Errors-in-variables model.

1. Summary

In the estimation problem for a multivariate errors-in-variables model, Gleser (1979) raised the question of nonsingularity of a principal submatrix of a random orthogonal matrix. This paper treats this problem in a wider framework. Let $\mathscr{S}(p)$ ($\mathscr{S}_+(p)$) be the set of $p\times p$ nonnegative (positive, respectively) definite matrices and let $S\in\mathscr{S}(p)$ be a $p\times p$ random matrix with a pdf $h(S|\theta)$ with respect to the Lebesgue measure $\mathrm{d}S$ on $\mathscr{S}(p)$, where θ is a point in an index set Θ. An example of S is the sample covariance matrix. Let $|S|$ denote the determinant of S and $\mathscr{O}(p)$ the set of $p\times p$ orthogonal matrices. Since the set $\{S\in\mathscr{S}(p): |S|=0\}$ has Lebesgue measure 0, the domain of $h(S|\theta)$ can be replaced by $\mathscr{S}_+(p)$. We decompose S as

$$S=GDG', \tag{1.1}$$

* Supported by the Ministry of Education of Japan under General C59540104.
** Supported by the Graduate School Research Committee, University of Wisconsin, Madison.

where $G \in \mathcal{O}(p)$ and

$$D = \begin{bmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_p \end{bmatrix} \quad (d_1 \geq \cdots \geq d_p).$$

Of course, the d_i's are the latent roots of S and they are distinct a.s. (see Okamoto (1973)).

In this situation, we would like to study the almost sure nonsingularity of a submatrix G_{11} of G where

$$G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \quad (G_{ij}: p_i \times p_j,\ p_1 + p_2 = p). \tag{1.2}$$

However, G is not uniquely determined even if the d_i's are distinct. In fact, if $G \in \mathcal{O}(p)$ satisfies (1.1), so does GE where

$$E \in \mathscr{E} = \left\{ E = \begin{bmatrix} \varepsilon_1 & & 0 \\ & \ddots & \\ 0 & & \varepsilon_p \end{bmatrix} : \varepsilon_i = +1 \text{ or } -1 \right\}.$$

Hence the problem considered here is whether or not the set

$$\mathcal{N} = \{[G] \in \mathcal{O}(p)/\mathscr{E} : G \in \mathcal{O}(p),\ |G_{11}| = 0\} \tag{1.3}$$

has measure 0 with respect to the induced measure over $\mathcal{O}(p)/\mathscr{E}$, where $\mathcal{O}(p)/\mathscr{E}$ denotes the left quotient space of $\mathcal{O}(p)$ by $\mathscr{E}$ and $[G] = G\mathscr{E}$ for $G \in \mathcal{O}(p)$. This paper gives a positive answer to this problem as follows.

Theorem. *The set $\mathcal{N}$ in* (1.3) *has measure* 0 *with respect to the measure on $\mathcal{O}(p)/\mathscr{E}$ induced by $h(S \mid \theta)\,\mathrm{d}S$.*

It is noted that for $G \in \mathcal{O}(p)$ and for $G^* \in [G]$, $|G_{11}| = 0$ if and only if $|G_{11}^*| = 0$, where G^* is partitioned as in (1.2). Hence, if a function of G is independent of the choice of $G^* \in [G]$ and depends on G_{11}^{-1}, it is defined almost surely with respect to the induced measure over $\mathcal{O}(p)/\mathscr{E}$.

In Section 3, the result is applied to Gleser's estimator. Here we remark that Gleser's proof of Lemma 2.2 in his paper (1979) is not complete since the choice of G in (1.1) is not unique.

2. Proof of the theorem

By excluding a set of Lebesgue measure 0, we regard the domain of $h(S \mid \theta)$ as the set

$$\tilde{S}_+(p) = \{S \in \mathscr{S}_+(p) : S \text{ has distinct roots}\}. \tag{2.1}$$

Let

$$\mathscr{D}=\left\{D=\begin{bmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_p \end{bmatrix} : d_1>d_2>\cdots>d_p>0\right\}.$$

Lemma 2.1. *$\bar{\mathscr{S}}_+(p)$ is homeomorphic to $\mathscr{O}(p)/\mathscr{E}\times\mathscr{D}$.*

Proof. Define the four maps

(1) f: $\mathscr{O}(p)\times\mathscr{D}\to\bar{\mathscr{S}}_+(p)$ by $f(G,D)=GDG'$,

(2) $\pi:\mathscr{O}(p)\times\mathscr{D}\to\mathscr{O}(p)/\mathscr{E}\times\mathscr{D}$ by $\pi(G,D)=([G],D)$,

(3) $\bar{f}$: $\mathscr{O}(p)/\mathscr{E}\times\mathscr{D}\to\bar{\mathscr{S}}_+(p)$ by $\bar{f}([G],D)=GDG'$, and

(4) pr: $\mathscr{O}(p)\to\mathscr{O}(p)/\mathscr{E}$ by $\mathrm{pr}(G)=[G]$.

Then $f=\bar{f}\circ\pi$. We shall show that $\bar{f}$ is a homeomorphism. The bijection of $\bar{f}$ is easily shown, and by $f=\bar{f}\circ\pi$, the continuity of $\bar{f}$ follows from that of f and the openness of the map π. To show the continuity of $\bar{f}^{-1}$, let $S_0\in\bar{\mathscr{S}}_+(p)$ and let $\{S_n\}$ be a sequence in $\bar{\mathscr{S}}_+(p)$ such that

$$\lim_{n\to\infty}\|S_n-S_0\|=0,$$

where $\|A\|=(\operatorname{tr}AA')^{1/2}$ for a matrix A. Write $\bar{f}^{-1}(S_j)=([G_j],D_j)$, where $S_j=G_jD_jG_j'$ $(j=0,n)$. First we show $\lim_{n\to\infty}\|D_n-D_0\|=0$. Since this is implied by

$$\|S_n-S_0\|\geq\|D_n-D_0\|, \tag{2.2}$$

we prove (2.2). Since $\operatorname{tr}S_nS_n'=\operatorname{tr}D_n^2$ and $\operatorname{tr}S_0S_0'=\operatorname{tr}D_0^2$, (2.2) is equivalent to $\operatorname{tr}S_nS_0'\leq\operatorname{tr}D_nD_0$, where the eigenvalues of D_n and D_0 are ordered. The last inequality follows from Mirsky (1975) or Wijsman (1979, Lemma 5.1). Next, to show $[G_n]\to[G_0]$ with respect to the induced topology, we suppose $[G_n]\not\to[G_0]$. Since $\mathscr{O}(p)/\mathscr{E}=\mathrm{pr}(\mathscr{O}(p))$ is compact Hausdorff and since it is easily shown to satisfy the first axiom of countability, it is sequentially compact. Hence there exists a subsequence $\{[G_{n'}]\}$ of $\{[G_n]\}$ such that $[G_{n'}]\to[G_1]\neq[G_0]$. But from

$$\|S_n-S_0\|=\|G_nD_nG_n'-G_0D_0G_0'\|\geq\big|\,\|D_n-D_0\|-\|G_nD_0G_n'-G_0D_0G_0'\|\,\big|$$

we obtain $\lim_{n\to\infty}\|G_nD_0G_n'-G_0D_0G_0'\|=0$. Since $\bar{f}([G_{n'}],D_0)=G_{n'}D_0G_{n'}'$ and $\bar{f}$ is continuous, this implies

$$\lim_{n'}\|G_{n'}D_0G_{n'}'-G_0D_0G_0'\|=\|G_1D_0G_1'-G_0D_0G_0'\|,$$

which in turn implies $[G_1]=[G_0]$. This is a contradiction and completes the proof.

Lemma 2.2. *Let ν be the invariant probability measure on $\mathscr{O}(p)$. Then $N=\{G\in\mathscr{O}(p): |G_{11}|=0\}$ has ν-measure 0.*

Proof. Let $X: p\times p\sim\mathrm{N}(0,I_p\otimes I_p)$. Since $|X|\neq0$ a.s., we assume $|X|\neq0$ and decompose X as $X=GW$ uniquely by Gram-Schmidt orthogonalization where

$G\in\mathcal{O}(p)$ and $W\in\mathcal{U}(p)=\{W: p\times p \mid W$ is upper triangular with positive diagonal elements$\}$. Then G has the distribution ν (see Eaton (1972), pp. 6.20, 6.21). Write

$$X=\begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22}\end{bmatrix} \quad\text{and}\quad W^{-1}=\begin{bmatrix} W^{11} & W^{12} \\ W^{21} & W^{22}\end{bmatrix}.$$

Then from $XW^{-1}=G$ and $W^{21}=0$, $G_{11}=X_{11}W^{11}$. Since $|W^{11}|\neq 0$, $|G_{11}|=0$ if and only if $|X_{11}|=0$. It is known that the set $\{X: |X_{11}|=0\}$ has measure 0 with respect to the multivariate normal measure. Since ν is induced by the distribution of X, N has ν-measure 0. This completes the proof.

Let $\mathrm{pr}: \mathcal{O}(p)\to\mathcal{O}(p)/\mathcal{E}$ be the projection map, i.e., $\mathrm{pr}(\Gamma)=[\Gamma]$, and let $\nu^*=\nu\circ\mathrm{pr}^{-1}$ be the induced measure over $\mathcal{O}(p)/\mathcal{E}$ by pr from ν, where ν is the invariant probability measure over $\mathcal{O}(p)$. Then ν^* is clearly invariant under the action of $\mathcal{O}(p)$ on $[G]\to\Gamma[G]$ for $\Gamma\in\mathcal{O}(p)$, and it is the unique invariant probability measure on $\mathcal{O}(p)/\mathcal{E}$ since the action of $\mathcal{O}(p)$ is transitive. Let $\bar{f}$ be the homeomorphism in the proof of Lemma 2.1 and let P be any probability distribution on $\mathcal{S}(p)$ such that P is equivalent to the Lebesgue measure $\mathrm{d}S$ and P is invariant under $S\to\Gamma S\Gamma'$ for $\Gamma\in\mathcal{O}(p)$. For example, P can be the Wishart distribution with mean nI_p and degrees of freedom n ($\geq p$). Then $Q=P\circ\bar{f}$ is the induced distribution of $([G],D)$. From the definition of $\bar{f}$, $Q(A)=0$ implies $P(B)=0$ and vice versa where $A=\bar{f}(B)$ with $A\subset\bar{\mathcal{S}}_+(p)$ and $B\subset\mathcal{O}(p)/\mathcal{E}\times\mathcal{D}$, and $\Gamma\bar{f}([G],D)\Gamma'=\bar{f}(\Gamma[G],D)$ for $\Gamma\in\mathcal{O}(p)$. Since P is invariant under $S\to\Gamma S\Gamma'$ and since the action $S\to\Gamma S\Gamma'$ on $\bar{\mathcal{S}}_+(p)$ corresponds to the action $([G],D)\to(\Gamma[G],D)$ on $\mathcal{O}(p)/\mathcal{E}\times\mathcal{D}$, Q is invariant under the action of $\mathcal{O}(p)$ on $\mathcal{O}(p)/\mathcal{E}\times\mathcal{D}$. Therefore by Theorem 3.4.1 in Farrel (1976), Q is factored as $Q=\nu^*\times\lambda$ where ν^* is the above unique invariant probability measure on $\mathcal{O}(p)/\mathcal{E}$ and λ is a regular Borel measure on $\mathcal{D}$. Consequently λ is a probability measure and for a Borel set C of $\mathcal{O}(p)/\mathcal{E}$,

$$P(\bar{f}(C,\mathcal{D}))=Q(C\times\mathcal{D})=\nu^*(C)\lambda(\mathcal{D})=\nu^*(C).$$

This implies that ν^* is the probability measure of $[G]$ induced from P. Thus for $\mathcal{N}$ in (1.3) and N in Lemma 2.2,

$$\nu^*(\mathcal{N})=\nu(\mathrm{pr}^{-1}(\mathcal{N}))=\nu(N)=0. \tag{2.3}$$

On the other hand, by assumption P is equivalent to $\mathrm{d}S$ and $\mathrm{d}S$ dominates $\mathrm{d}P_\theta\equiv h(S\mid\theta)\,\mathrm{d}S$. Hence the induced measure of $[G]$ on $\mathcal{O}(p)/\mathcal{E}$ under $h(S\mid\theta)\,\mathrm{d}S$ is dominated by ν^*. In fact, $\nu^*(C)=P(\bar{f}(C,\mathcal{D}))=0$ implies $P_\theta(\bar{f}(C,\mathcal{D}))=0$. Therefore the proof of the Theorem is now completed.

3. Application

Let $x_1,\ldots,x_n$ be independent random $p\times 1$ vectors and let each x_i have a pdf $f_i(x_i-\mu_i)$ with respect to the Lebesgue measure on R^p, where μ_i is a $p\times 1$ vector.

Then the random matrix $S=X'X=\sum_{i=1}^{n} x_i x_i'$, where $X'=[x_1,x_2,\ldots,x_n]$, has a pdf $h(S\mid\theta)$ with respect to the Lebesgue measure on $\mathscr{S}(p)$. Here $\theta=(\mu_1,\ldots,\mu_n)$. The assumptions of Section 1 are satisfied and the conclusion of the theorem holds.

In particular this can be applied to the situation Gleser (1979) considered. In estimating or testing the linear functional relationship for the multivariate errors-in-variables model, he has derived an estimator which depends on the random matrix $G_{21}G_{11}^{-1}$ via the decomposition (1.1). Since $G_{21}G_{11}^{-1}$ can be regarded as a function of $[G]$ and it is independent of the choice of G^* in $[G]$, our theorem guarantees the a.s. existence of $G_{21}G_{11}^{-1}$. In fact, he assumed that the f_i's are multivariate normal with covariance matrix $\sigma^2 I_p$, in which case the estimator becomes the maximum likelihood estimator. The reader may consult his paper for details.

Another application is found in the field of distribution theory. Sometimes the distribution of the column vectors of G or the latent vectors of S in (1.1) is considered with the sign restrictions on some elements of each vector. For example, when the i-th latent vector is treated, it is often the case that the i-th element of the vector is assumed to be positive. However, this can be done only when the i-th element is nonzero a.s. According to our theorem, it is nonzero a.s. with respect to the measure on $\mathscr{O}(p)/\mathscr{E}$ induced by $h(S\mid\theta)\,\mathrm{d}S$, but not with respect to $h(S\mid\theta)\,\mathrm{d}S$ itself.

Acknowledgement

The authors are very grateful to Professor R.A. Wijsman for pointing out the non-triviality of the bicontinuity of $\bar{f}$ in Lemma 2.1 and to the referee for helpful comments.

References

Eaton, M.L. (1972). *Multivariate Analysis.* Univ. of Copenhagen, Copenhagen.

Eaton, M.L. and Perlman, M.D. (1973). The nonsingularity of generalized sample covariance matrices. *Ann. Statist.* 1, 710–717.

Farrel, R.H. (1976). *The Technique of Multivariate Calculation.* Springer, Berlin.

Gleser, L.J. (1979). Estimation in multivariate 'errors in variables' regression model: large sample results. *Ann. Statist.* 9, 24–44.

Mirsky, L. (1975). A trace inequality of John von Neumann. *Monatsh. Math.* 79, 303–305.

Okamoto, M. (1973). Distinctness of the eigenvalues of a quadratic form in multivariate sample. *Ann. Statist.* 1, 763–765.

Wijsman, R.A. (1979). Constructing all smallest simultaneous confidence sets in a given class, with application to MANOVA. *Ann. Statist.* 7, 1003–1013.