S. A. Andersson M. D. Perlman

Lattice Models for Conditional Independence in a Multivariate Normal Distribution

Preprint
December
1988
9

Institute of Mathematical Statistics
University of Copenhagen

ISSN 0902-8846

S.A. Andersson and M.D. Perlman[* **]

LATTICE MODELS FOR CONDITIONAL INDEPENDENCE IN A MULTIVARIATE NORMAL DISTRIBUTION

Preprint 1988 No. 9

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

December 1988

*) Department of Statistics, University of Washington

**) This research was supported in part by the Danish Research Council and by U.S. National Science Foundation Grant No. DMS 86-03489

Summary

Let $\mathcal{K}$ be a distributive lattice of quotient spaces of a vector space V and for $K \in \mathcal{K}$ let $p_K: V \to K$ be the projection onto K. The statistical model $\mathcal{N}_V(\mathcal{K})$ is defined to be the set of all normal distributions on V such that for every pair $L, M \in \mathcal{K}$, p_L and p_M are conditionally independent given $p_{L \wedge M}$. Statistical properties of $\mathcal{N}_V(\mathcal{K})$ are studied, eg., maximum likelihood inference, invariance, and the problem of testing H_0: $\mathcal{N}_V(\mathcal{K})$ vs H: $\mathcal{N}_V(\mathcal{M})$ when $\mathcal{M}$ is a distributive sublattice of $\mathcal{K}$. The set $J(\mathcal{K})$ of join-irreducible elements of $\mathcal{K}$ plays a central role in the analysis of $\mathcal{N}_V(\mathcal{K})$. This class of statistical models appears to generate all multivariate normal conditional independence models for which complete and explicit likelihood inference is possible.

AMS 1980 subject classification: Primary 62H12, 62H15; Secondary 62H20, 62H25.

Key words and phrases: Distributive lattice, quotient spaces, join-irreducible elements, pairwise conditional independence, multivariate normal distribution, invariance, generalized block-triangular matrices, maximum likelihood estimator, likelihood ratio statistic.

CONTENTS

§1. INTRODUCTION.

In recent years the study of conditional independence (CI) models in multivariate normal distributions has received increasing attention. Prominent references include Dempster (1972), Kiiveri, Speed, and Carlin (1984), Lauritzen (1985, 1989), Lauritzen, Dawid, Larsen and Leimer (1988), Lauritzen and Frydenberg (1988), Lauritzen and Wermuth (1984, 1987), Speed and Kiiveri (1986), and Wermuth (1976, 1980, 1985, 1988). In most of these studies the CI assumptions are equivalent to the occurrence of certain patterns of zeroes in the precision matrix Σ^{-1} of a multivariate normal distribution, hence the models are linear in Σ^{-1}. These conditions are similar to those occurring in certain multiplicative or log-linear models for categorical data - cf. A. H. Andersen (1974), E.B. Andersen (1980), Darroch, Lauritzen, and Speed (1980), Darroch and Speed (1983), Goodman (1970, 1971), Haberman (1974), Sundberg (1975), Wermuth and Lauritzen (1983) and many others. Dawid (1979, 1980) has presented a useful summary of general properties of CI.

An interesting general class of CI models is the class of multivariate graphical chain models recently introduced by Lauritzen and Wermuth (1987). Although this class of models may have extensive applications, it appears to be too large to permit a complete and explicit statistical analysis, eg., explicit likelihood inference.

In the present paper we shall define and study a somewhat smaller but still very rich class of CI models, namely those determined by the requirement of pairwise CI with respect to a distributive lattice $\mathcal{K}$ of quotient spaces of the observation space V. Under the added assumption of multivariate normality, every such statistical model admits a complete and explicit likelihood analysis which parallels that in classical multivariate analysis where Σ is unrestricted. This class of models includes many if not all of the CI models that have been studied in detail in multivariate analysis, eg., Das Gupta (1977), Banerjee and Giri (1980), Marden (1981), and appears to generate all CI models for which complete and explicit likelihood inference is possible.

This class of models arose in the following way. Many investigators have realized that in a balanced ANOVA design, the lattice structure of the collection of <u>linear subspaces</u> that comprise the design plays a fundamental role in its analysis. In fact, the design is balanced if and only if this lattice is <u>distributive</u>. In this case Andersson (1987) explicitly demonstrated the central role played by the poset of <u>join-irreducible elements</u> of this lattice for the analysis of the design, eg., for the construction of the ANOVA table.

During the course of Andersson's investigation it was realized that the dual notion of a <u>distributive lattice of quotient spaces</u> may be applied in a natural way to define a rich class of CI models that is amenable to explicit analysis. This analysis is facilitated by the fact that each such model is invariant under a group of generalized block-triangular linear transformations A that acts transitively on the model.

These models are defined as follows. Let V be a real finite-dimensional vector space and let $\mathcal{K}$ be a distributive lattice of quotient spaces of V (cf. Section 2). For each $K \in \mathcal{K}$ let $p_K: V \to K$ denote the projection onto K. Then the model $\mathcal{N}_V(\mathcal{K})$ is defined to be the set of all normal distributions N on V such that for every pair $L, M \in \mathcal{K}$, p_L and p_M are CI given

$p_{L\Lambda M}$. (For simplicity we shall always assume that each N has mean vector 0 and nonsingular covariance matrix denoted by Σ).

For example, suppose that $V = \mathbb{R}^{\{1,2,3\}}$ and that

$$(1.1) \qquad \mathcal{K} = \{\mathbb{R}^{\{1\}}, \mathbb{R}^{\{1,2\}}, \mathbb{R}^{\{1,3\}}, \mathbb{R}^{\{1,2,3\}}\}.$$

Then $\mathcal{N}_V(\mathcal{K})$ consists of all normal distributions on V such that (x_1, x_2) and (x_1, x_3) are CI given x_1, i.e., x_2 and x_3 are CI given x_1. Equivalently, the precision matrix Σ^{-1} must satisfy the condition $(\Sigma^{-1})_{23} = (\Sigma^{-1})_{32} = 0$, so in this example the model is linear in Σ^{-1}. (This need not hold in the case of a general lattice $\mathcal{K}$, however - cf. Examples 3.6, 3.7, and 3.8 in Section 3.) This model is invariant under nonsingular linear transformations of $\mathbb{R}^{\{1,2,3\}}$ of the form

$$x \equiv \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \to \begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & 0 & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \equiv Ax.$$

It is easily verified that the set of all such nonsingular matrices A forms a subgroup of the group of all 3×3 lower triangular matrices.

It will be seen in Section 4 that for a general distributive lattice $\mathcal{K}$, the partitioning of the matrix A and the location of its zeroes is determined by $J(\mathcal{K})$, the poset of join-irreducible elements of $\mathcal{K}$. As in the case of a balanced ANOVA design, this poset completely determines the structure and analysis of the model $\mathcal{N}_V(\mathcal{K})$.

This paper is organized as follows. Section 2 contains the basic Decomposition Theorem (Theorem 2.1) which shows that the observation space V can be represented as a product of vector spaces indexed by $J(\mathcal{K})$ in such a way that for each $K \in \mathcal{K}$, the projection $p_K: V \to K$ becomes simply a coordinate projection (cf. Remark 2.1). By means of this representation one may choose a $\mathcal{K}$-adapted basis for V (cf. Remark 2.2) which is later shown to yield natural matrix representations for the model $\mathcal{N}_V(\mathcal{K})$ and its invariance group.

In Theorem 3.1 it is shown that $N \in \mathcal{N}_V(\mathcal{K})$ if and only if the members of $\mathcal{K}$ are geometrically orthogonal (GO) with respect to the inner product determined by the precision operator δ of N. This reduces the statistical problem of characterizing the distributions in $\mathcal{N}_V(\mathcal{K})$ to the algebraic problem of determining those positive definite forms δ which render the members of $\mathcal{K}$ GO.

Thus the model $\mathcal{N}_V(\mathcal{K})$ may be parametrized either by a constrained set of precision operators δ on V or, equivalently, by the corresponding constrained set of covariance operators $\sigma = \delta^{-1}$ on the dual space of V. Once a $\mathcal{K}$-adapted basis for V has been selected, the model may be parametrized either by a constrained set of precision matrices Σ^{-1} or by the corresponding constrained set of covariance matrices Σ. It is shown in Sections 3.2 and 3.3, however, that these parameter sets can be represented as products of smaller unconstrained parameter sets, again indexed by $J(\mathcal{K})$, from which the maximum likelihood (ML) estimators and the likelihood ratio (LR) test statistics may be readily determined. These alternative parametrizations are called the $\mathcal{K}$-parametrizations of the model $\mathcal{N}_V(\mathcal{K})$.

For example, if $\mathcal{K} = \{K,V\}$ is a simple chain ($K < V$) and a $\mathcal{K}$-adapted basis for V is chosen, then Σ is unconstrained and its $\mathcal{K}$-parametrization under $\mathcal{N}_V(\mathcal{K})$ reduces to the well-known 1-1 correspondence

$$\Sigma \longleftrightarrow (\Sigma_{11},\ \Sigma_{21}\Sigma_{11}^{-1},\ \Sigma_{22\cdot 1})$$

When $\mathcal{K}$ is given by (1.1), however, the CI constraints imposed upon Σ are non-trivial and its $\mathcal{K}$-parametrization under $\mathcal{N}_V(\mathcal{K})$ is given by

$$\Sigma \longleftrightarrow (\Sigma_{11},\ \Sigma_{21}\Sigma_{11}^{-1},\ \Sigma_{22\cdot 1},\ \Sigma_{31}\Sigma_{11}^{-1},\ \Sigma_{33\cdot 1})$$

(cf. Example 3.6). The ranges of the components of these $\mathcal{K}$-parametrizations are <u>unconstrained</u> (except for the trivial requirement that Σ_{11}, $\Sigma_{22\cdot 1}$, and $\Sigma_{33\cdot 1}$ be positive definite).

A step-wise algorithm for reconstructing Σ from its $\mathcal{K}$-parameters is presented in Remark 3.6.

The group of generalized block-triangular transformations A that leave the model $\mathcal{N}_V(\mathcal{K})$ invariant is defined and studied in Section 4. In particular, it is shown that this group acts <u>transitively</u> on the model.

The normal statistical model $\mathcal{N}_V(\mathcal{K})$ is formally defined in Section 5. The likelihood function is decomposed as a product of conditional densities involving only the $\mathcal{K}$-parameters of Σ (cf. (5.5)), from which the maximum likelihood estimators of the $\mathcal{K}$-parameters are easily derived and used in turn to obtain the ML estimator $\hat{\Sigma}$ by means of the step-wise reconstruction algorithm. In Remark 5.4 it is noted that the model $\mathcal{N}_V(\mathcal{K})$ is determined by a system of <u>linear recursive equations</u> with lattice structure (cf. Wermuth (1980), Kiiveri, Speed, Carlin (1984)).

Section 6 treats the problem of testing one such model against another, eg., testing

(1.2) $$H_0: \mathcal{N}_V(\mathcal{K}) \quad \text{vs.} \quad H: \mathcal{N}_V(\mathcal{M})$$

where $\mathcal{M}$ is a distributive sublattice of $\mathcal{K}$. (Note that $\mathcal{M} \subset \mathcal{K} \Rightarrow \mathcal{N}_V(\mathcal{K}) \subseteq \mathcal{N}_V(\mathcal{M})$.) The LR statistic Q is easily expressed in terms of the ML estimates of the $\mathcal{K}$-parameters and $\mathcal{M}$-parameters of Σ. The central distribution of Q is then derived by means of the invariance of the testing problem.

For example, if $\mathcal{K}$ is given by (1.1) and $\mathcal{M} = \{\mathbb{R}^{\{1,2,3\}}\}$, then $\mathcal{N}_V(\mathcal{M})$ is the unrestricted normal model and (1.2) becomes the problem of testing that x_2 and x_3 are CI given x_1, i.e., testing

(1.3) $$H_0: \sigma_{23} = \sigma_{21}\sigma_{11}^{-1}\sigma_{13} \quad \text{vs.} \quad H: \sigma_{23} \neq \sigma_{21}\sigma_{11}^{-1}\sigma_{12},$$

where $\Sigma = (\sigma_{ij} | i,j = 1,2,3)$. If, however, $\mathcal{M}$ is given by (1.1) while

(1.4) $$\mathcal{K} = \{\{0\}, \mathbb{R}^{\{1\}}, \mathbb{R}^{\{3\}}, \mathbb{R}^{\{1,2\}}, \mathbb{R}^{\{1,3\}}, \mathbb{R}^{\{1,2,3\}}\},$$

then (1.2) becomes the problem of testing

(1.5) $$H_0: \sigma_{13} = \sigma_{23} = 0 \quad \text{vs.} \quad H: \sigma_{23} = \sigma_{21}\sigma_{11}^{-1}\sigma_{13}.$$

Notice that when $\mathcal{K}$ is given by (1.4) then $\mathcal{N}_V(\mathcal{K}) = \mathcal{N}_V(\mathcal{K}')$, where

$$\mathcal{K}' = \{\{0\}, \mathbb{R}^{\{1,2\}}, \mathbb{R}^{\{3\}}, \mathbb{R}^{\{1,2,3\}}\} \subset \mathcal{K},$$

and that this model is equivalent to the hypothesis that (x_1,x_2) and x_3 are independent. Thus two different lattices $\mathcal{K}$, $\mathcal{K}'$ may determine the same

CI model. The question of characterizing a minimal determining lattice for a given CI model is currently under study.

A series of examples is presented in Section 3 and continued in Sections 4-6 to illustrate these concepts and results. Several possible extensions of this class of lattice models for CI are suggested in Section 7.

Finally, it is important to mention that the CI models $\mathcal{N}_V(\mathcal{K})$ play an important role in the analysis of <u>non-nested</u> missing data models. Under the assumption of multivariate normality it is well known that a <u>nested</u> missing data model with <u>unrestricted</u> covariance matrix Σ admits a complete and explicit likelihood analysis, remaining invariant under the appropriate group of block-triangular transformations, which acts transitively on the unrestricted set of covariance matrices (cf. Eaton and Kariya (1983), Andersson, Marden, and Perlman (1988)). If the missing data pattern is non-nested, however, then explicit analysis is not possible in general.

If, however, the missing data pattern is non-nested but is determined by a <u>distributive lattice</u> $\mathcal{K}$ of <u>quotient spaces of</u> V, then a complete and explicit likelihood analysis is possible under the additional assumption that $\mathcal{N}_V(\mathcal{K})$ holds, or, more precisely, that the covariance structure Σ satisfies the CI restriction determined by $\mathcal{K}$ (cf. Andersson, Marden, and Perlman (1989)). (The theory developed in the present paper allows one to test the validity of this assumption.) Under this assumption, the non-nested missing data model determined by $\mathcal{K}$ remains invariant under the group of generalized block-triangular transformations A studied in the present paper, which group acts transitively on the restricted set of covariance matrices that determine the model $\mathcal{N}_V(\mathcal{K})$. Thus, the results in the present paper open the possibility of applying classical multivariate techniques to a class of missing data models much larger than the nested class.

Throughout this paper frequent reference will be made to the papers Andersson (1982) and Andersson (1987) - these will be abbreviated as [A] (1982) and [A] (1987), respectively.

§2. THE LATTICE OF QUOTIENT SPACES OF A VECTOR SPACE.

2.1. The lattice structure of quotient spaces.

Let V be a finite dimensional real vector space with zero element 0. A quotient space (or simply a quotient) of V is formally defined to be a pair (K,p) consisting of a vector space K and a surjective linear mapping $p:V \to K$. For ease of notation, (K,p) usually is abbreviated to K. Sometimes the mapping p is written as p_{KV} or simply p_K when V is understood.

Let L and M be two quotients of V. If there exists a linear mapping $p_{LM}:M \to L$ such that $p_L = p_{LM} \circ p_M$ then p_{LM} is necessarily surjective and unique, hence (L,p_{LM}) is a quotient of M. In this situation we write $(L,p_L) \leq (M,p_M)$, or simply $L \leq M$. This relation is equivalent to the condition that $p_L^{-1}(0) \supseteq p_M^{-1}(0)$. The relation $\leq$ on the set of all quotients of V is not antisymmetric, hence one defines an equivalece relation $\sim$ on this set by $L \sim M$ if $p_L^{-1}(0) = p_M^{-1}(0)$. The collection of equivalence classes is denoted by $\mathcal{K}(V)$. Equipped with the relation induced by $\leq$ (also denoted by $\leq$), $\mathcal{K}(V)$ becomes a partially ordered set ($\equiv$ poset - cf. [A] (1987), Section 1.1).

We identify a quotient (K,p_K) of V with its equivalence class in $\mathcal{K}(V)$. A convenient representive for this equivalence class is the canonical quotient space $(V/p_K^{-1}(0),p)$, where $p:V \to V/p_K^{-1}(0)$ is the canonical mapping given by $p(x) = x+p_K^{-1}(0)$, $x \in V$.

The poset $\mathcal{K}(V)$ is in fact a lattice: if $L,M \in \mathcal{K}(V)$ then their minimum and maximum exist and are given by

$$L \wedge M := V/(p_L^{-1}(0)+p_M^{-1}(0))$$
$$L \vee M := V/(p_L^{-1}(0)\cap p_M^{-1}(0))$$

respectively (cf. [A] (1987), Section 1.3.). Since V is finite dimensional, the lattice $\mathcal{K}(V)$ has finite length. The minimal and maximal elements exist and are given by $\{0\}$ and V respectively. If $\dim(V) \geq 2$ then $\mathcal{K}(V)$ is not distributive and $|\mathcal{K}(V)| = \infty$ (cf. [A] (1987), Section 1.4.).

2.2. Distributive lattices of quotient spaces.

As stated in the Introduction, the covariance models studied in this paper are those determined by pairwise CI with respect to a distributive sublattice $\mathcal{K} \subseteq \mathcal{K}(V)$ of quotient spaces of V. Since $\mathcal{K}(V)$ and therefore also $\mathcal{K}$ is of finite length, if $\mathcal{K}$ is distributive it must be finite (cf. [A] (1987), Proposition 1.1). A particular class of distributive sublattices $\mathcal{K} \subseteq \mathcal{K}(\mathbb{R}^I)$ is described in the following example.

Example 2.1. For a finite index set I, let $\mathfrak{R}$ be a ring of subsets of I. For each $R \in \mathfrak{R}$ define the mapping $p_R:\mathbb{R}^I \to \mathbb{R}^R$ by $p_R((x_i|i\in I)) = (x_i|i\in R)$.

Since $\mathcal{R}$ is a ring, it follows that $\mathcal{K}(\mathcal{R}) := \{(\mathbb{R}^R, p_R) | R \in \mathcal{R}\}$ is a distributive lattice of quotients of the vector space $\mathbb{R}^I =: V$. If $I \in \mathcal{R}$, then $\mathbb{R}^I \in \mathcal{K}(\mathcal{R})$. Thus, by the classical theorem of Birkhoff and Stone, every finite distributive lattice is isomorphic to a distributive lattice of quotients of a finite-dimensional vector space. □

It will be shown in Remark 2.2 that in fact <u>every</u> distributive sublattice $\mathcal{K} \subseteq \mathcal{K}(V)$ can be represented in the form $\mathcal{K} = \mathcal{K}(\mathcal{R})$ for some $\mathcal{R}$ as in Example 2.1.

Let $J(\mathcal{K})$ denote the poset of all join-irreducible elements in $\mathcal{K}$, i.e.,

$$J(\mathcal{K}) = \{K \in \mathcal{K} | \ \vee(K' \in \mathcal{K} | K' < K) < K\} \cup \{0_{\mathcal{K}}\}$$

where $0_{\mathcal{K}}$ denotes the minimal element in $\mathcal{K}$ (cf. [A] (1987), Section 1.5). If $K \in J(\mathcal{K})$ and $K \neq 0_{\mathcal{K}}$, then define

$$J(K) := \vee(K' \in \mathcal{K} | K' < K) < K;$$

also, define $J(0_{\mathcal{K}}) = \{0\}$. Note that $J(0_{\mathcal{K}})$ need not be an element in $\mathcal{K}$.

In the following theorem the space V is represented as a product of vector spaces indexed by the poset $J(\mathcal{K})$ of join-irreducible elements of $\mathcal{K}$, such that the space with index $K \in J(\mathcal{K})$ has dimension $\dim(K) - \dim(J(K))$. This decomposition is applied in Section 3 to characterize the structure of the covariance model determined by pairwise CI with respect to $\mathcal{K}$. Thus the reader should be aware of the fundamental role of the poset $J(\mathcal{K})$ in this theory.

<u>Theorem 2.1.</u> (Decomposition Theorem). Let $\mathcal{K} \subseteq \mathcal{K}(V)$ be a distributive lattice such that $V \in \mathcal{K}$. For each $K \in J(\mathcal{K})$, let $r_K : K \to p^{-1}_{J(K),K}(0)$ be any surjective linear mapping. Then the linear mapping

$$(2.1) \qquad \begin{aligned} \varphi_V : V &\to X(p^{-1}_{J(K),K}(0) \mid K \in J(\mathcal{K})) \\ x &\to (r_K(p_K(x)) \mid K \in J(\mathcal{K})) \end{aligned}$$

is bijective.

<u>Proof</u>: For $K \in \mathcal{K}$ the set $\mathcal{K}_K = \{K' \in \mathcal{K} \mid K' \leq K\}$ is a distributive lattice of quotients of K with K itself as the maximal element and $J(\mathcal{K}_K) = J(\mathcal{K}) \cap \mathcal{K}_K$. We shall prove the theorem by using induction on the cardinality $|J(\mathcal{K})|$ of $J(\mathcal{K})$. If $|J(\mathcal{K})| = 1$ or 2, the result is trivial. Suppose that the result is true when $|J(\mathcal{K})| \leq n-1$, and assume that $|J(\mathcal{K})| = n$. First, if $V \in J(\mathcal{K})$ then the mapping

$$\begin{aligned} J(V) &\to X(\, p^{-1}_{J(K),K}(0) \mid K \in J(\mathcal{K}_{J(V)})) \\ x &\to (r_K(p_{K,J(V)}(x)) \mid K \in J(\mathcal{K}_{J(V)})) \end{aligned}$$

is bijective because $|J(\mathcal{K}_{J(V)})| = n-1$. Since the linear mapping

$$\begin{aligned} V &\to J(V) \times p^{-1}_{J(V)}(0) \\ x &\to (p_{J(V)}(x), r_V(x)) \end{aligned}$$

is bijective and $p_{K,J(V)} \circ p_{J(V)} = p_K$ for every $K \in J(\mathcal{K}_{J(V)})$, the mapping (2.1) is bijective in this case.

If, on the other hand, $V \notin J(\mathcal{K})$, i.e., $V = L \vee M$, where $L < V$ and $M < V$, then it follows from [A] (1987) (Lemma 1.2 and the proof of Theorem 2.1) that $|J(\mathcal{K}_L)| < n$ and $|J(\mathcal{K}_M)| < n$. By the induction assumption it then follows that the mapping

$$\begin{aligned} V &\to X(p^{-1}_{J(K),K}(0)\,|K\in J(\mathcal{K}_L)) \\ x &\to \quad (r_K(p_K(x))\,|K\in J(\mathcal{K}_L)) \end{aligned}$$

is (equivalent to) the quotient L in V. Furthermore, M and $L \wedge M$ can be represented in an analogous way. Therefore,

$$\begin{aligned} p^{-1}_{L\wedge M,L}(0) &= X(p^{-1}_{J(K),K}(0)\,|K\in J(\mathcal{K}_L),K\nleq L\wedge M) \\ p^{-1}_{L\wedge M,M}(0) &= X(p^{-1}_{J(K),K}(0)\,|K\in J(\mathcal{K}_M),K\nleq L\wedge M). \end{aligned}$$

By Lemma 2.1 below, the linear mapping

$$\begin{gathered} \varphi_V : V \to \\ X(p^{-1}_{J(K),K}(0)\,|K\in J(\mathcal{K}_{L\wedge M})) \times X(p^{-1}_{J(K),K}(0)\,|K\in J(\mathcal{K}_L),K\nleq L\wedge M) \\ \times X(p^{-1}_{J(K),K}(0)\,|K\in J(\mathcal{K}_M),\ K\nleq L\wedge M) \\ = X(p^{-1}_{J(K),K}(0)\,|K\in J(\mathcal{K})). \end{gathered}$$

is bijective. Note that the above equality follows from the relations $J(\mathcal{K}) = J(\mathcal{K}_L) \cup J(\mathcal{K}_M)$ and $J(\mathcal{K}_{L\wedge M}) = J(\mathcal{K}_L) \cap J(\mathcal{K}_M)$. □

If $V \notin \mathcal{K}$, the theorem may be applied to the extended lattice $\mathcal{K} \cup \{V\}$ (also distributive). Thus, for the remainder of this paper it is assumed that $V \in \mathcal{K}$.

Remark 2.1. The representation (2.1) shows that V can be identified with a product of vector spaces indexed by $J(\mathcal{K})$; similary, each $L \in \mathcal{K}$ can be identified with the product $X(p^{-1}_{J(K),K}(0)\,|\,K\in J(\mathcal{K}),\ K\leq L)$ through the bijective linear mapping φ_L defined by $\varphi_L(x) = (r_K(p_K(x))\,|\,K\in J(\mathcal{K}),\ K\leq L)$, $x \in L$; Under these identifications, each mapping p_{LM}, $L \leq M \leq V$, is simply a canonical projection mapping. □

Remark 2.2. For each $K \in J(\mathcal{K})$, let D_K be a set with

$$|D_K| = \dim(p^{-1}_{J(K),K}(0)).$$

For $L \in \mathcal{K}$, define

$$I_L = \dot{\bigcup}(D_K\,|\,K\in J(\mathcal{K}),\ K\leq L) \tag{2.2}$$

and define $I = I_V$. The set $\mathcal{R}(\mathcal{K}) := \{I_L\,|\,L\in\mathcal{K}\}$ of subsets of I is a ring, i.e., a distributive lattice under the usual operations $\cup$ and $\cap$, and it is isomorphic to $\mathcal{K}$ through the lattice isomorphism $L \to I_L$, $L \in \mathcal{K}$. From Remark 2.1 it follows that there exists a basis $(e_i\,|\,i\in I)$ for V such that the elements (K,p_K) in $\mathcal{K}$ can be represented as

$$K = \operatorname{span}\{e_i\,|\,i\in I_K\}, \tag{2.3}$$

$$p_K(e_i) = \begin{cases} e_i & \text{for } i\in I_K \\ 0 & \text{for } i\in I\setminus I_K. \end{cases}$$

We shall say that a basis with the property (2.3) is <u>adapted</u> to $\mathcal{K}$. Thus, when V is identified with $\mathbb{R}^I$ through an adapted basis $(e_i|i\in I)$, the distributive lattice $\mathcal{K} \subseteq \mathcal{K}(V)$ is identified with $\mathcal{K}(\mathcal{R}(\mathcal{K}))$ (cf. Example 2.1).□

Remark 2.3. For $K \in J(\mathcal{K})$, note that

$$I_K = I_{J(K)} \dot{\cup} D_K \tag{2.4}$$

and

$$p^{-1}_{J(K),K}(0) = \mathrm{span}\{e_i | i\in D_K\}. \tag{2.5}$$

In particular, if $K = 0_{\mathcal{K}}$ then $J(K) = \{0\}$, $I_{J(K)} = \emptyset$, and $I_K = D_K$. □

Remark 2.4. From the isomorphism φ_L in Remark 1.1 one obtains directly that

$$\dim(L) = \Sigma(\dim(K)-\dim(J(K)) | K\in J(\mathcal{K}),\ K\leq L), \qquad L \in \mathcal{K}; \tag{2.6}$$

equivalently,

$$\dim(K)-\dim(J(K)) = \Sigma(\mu(K,L)\dim(L) | L\in J(\mathcal{K})), \qquad L \in J(\mathcal{K}), \tag{2.7}$$

where μ is the Möbius function for the poset $J(\mathcal{K})$ (cf. [A] (1987), Lemma 1.1). □

The following lemma is needed to complete the proof of Theorem 2.1.

Lemma 2.1. Let $L, M \in \mathcal{K}(V)$ with $L \vee M = V$ and let $r_L : L \to p^{-1}_{L\wedge M,L}(0)$ and $r_M : M \to p^{-1}_{L\wedge M,M}(0)$ be surjective linear mappings. Then the linear mapping

$$(2.8) \qquad \begin{aligned} \varphi : V &\to (L\wedge M) \times p^{-1}_{L\wedge M,L}(0) \times p^{-1}_{L\wedge M,M}(0) \\ x &\to (p_{L\wedge M}(x), r_L(p_L(x)), r_M(p_M(x))) \end{aligned}$$

is bijective.

Proof: Suppose that $\varphi(x) = 0$. Then $p_{L\wedge M}(x) = 0$ and we obtain that $p_L(x) \in p^{-1}_{L\wedge M,L}(0)$. In fact $p_L(x) = 0$ since r_L is surjective. Similarly $p_M(x) = 0$, hence $x \in p_L^{-1}(0) \cap p_M^{-1}(0) = 0$. The linear mapping φ is thus injective and a simple dimension argument shows that φ is also surjective. □

Remark 2.5. The lattice $\mathcal{K}(V)$ of quotients of V is isomorphic to the dual (cf. Grätzer (1978), pp.2-6) of the lattice $\mathcal{L}(V)$ of subspaces of V studied in [A] (1987) under the correspondence

$$(2.9) \qquad \begin{aligned} \mathcal{K}(V) &\longleftrightarrow \mathcal{L}(V) \\ K &\to p_K^{-1}(0) \\ V/L &\leftarrow L. \end{aligned}$$

Note that the operations $\vee$, $\wedge$ in $\mathcal{K}(V)$ correspond to $\cap$, $+$ in $\mathcal{L}(V)$. Thus by the Duality Principle (Grätzer (1978), p 6), all results for $\mathcal{K}(V)$ may be obtained from corresponding results for $\mathcal{L}(V)$. For example, Theorem 2.1 above may be obtained from an appropriate reformulation of Theorem 2.1 of [A] (1987). □

2.3 Geometrically orthogonal lattices of quotient spaces.

Let δ be an inner product on V and let (K,p_K) be a quotient of V. Let $p_K^{-1}(0)^{\perp}$ denote the orthogonal complement of $p_K^{-1}(0)$ with respect to (wrt) δ and let $q_K:V \to V$ be the orthogonal projection onto $p_K^{-1}(0)^{\perp}$ wrt δ.

Definition 2.1. Two quotients L and M of V are said to be geometrically orthogonal (GO) wrt δ if the orthogonal projections q_L and q_M commute. □

Remark 2.6. It is seen from Definition 2.1 of [A] (1987) that L and M are GO quotient spaces wrt δ if and only if $p_L^{-1}(0)$ and $p_M^{-1}(0)$ are GO subspaces wrt δ. □

Theorem 2.2. Let δ be an inner product on V and let $\mathcal{K} \subseteq \mathcal{K}(V)$ be a lattice of quotients of V. If the elements in $\mathcal{K}$ are geometrically orthogonal wrt δ then $\mathcal{K}$ is distributive. Conversely, if $\mathcal{K} \subseteq \mathcal{K}(V)$ is a distributive lattice then there exists an inner product δ such that all elements in $\mathcal{K}$ are geometrically orthogonal wrt δ.

Proof: Since $\mathcal{L} = \{p_K^{-1}(0) \mid K \in \mathcal{K}\}$ is a lattice of GO subspaces in V, it follows from Proposition 2.1 of [A] (1987) that $\mathcal{L}$ is distributive. By the duality between $\mathcal{L}(V)$ and $\mathcal{K}(V)$ (cf. Remark 2.5), it follows that $\mathcal{K}$ is distributive. Conversely, if $\mathcal{K}$ is distributive, then also $\mathcal{L}$ is distributive and it follows from Proposition 2.2 of [A] (1987) that there exists an inner product δ on V such that all elements in $\mathcal{L}$ are GO wrt δ. By Remark 1.6, all elements in $\mathcal{K}$ are GO wrt δ. □

§3. COVARIANCE STRUCTURE DETERMINED BY PAIRWISE CONDITIONAL INDEPENDENCE WITH RESPECT TO A DISTRIBUTIVE LATTICE OF QUOTIENT SPACES.

In this section the covariance structure determined by pairwise conditional independence (CI) with respect to a distributive lattice of quotient spaces is characterized in a multivariate normal distribution.

3.1. Algebraic conditions for conditional independence.

Let V be a real finite-dimensional vector space, δ an inner product ($\equiv$ positive definite form) on V, and N_δ the multivariate normal distribution on V with mean 0 and precision[1] δ. Since $\delta \in \mathbf{P}(V)$:= the set of all positive definite forms on V, its inverse $\delta^{-1} \equiv \sigma \in \mathbf{P}(V^*)$, where V^* is the dual space of V. When δ is a positive definite form on V, we abbreviate $\delta(x,x)$ to $\delta(x)$ for $x \in V$.

Let (K,p_K) be a quotient of V, $p_K^t:K^* \to V^*$ the dual mapping of p_K, and δ_K the precision of the transformed normal distribution $p_K(N_\delta)$ on K. Then $\delta_K = (\delta^{-1}\circ(p_K^t \times p_K^t))^{-1}$, since $\delta^{-1}\circ(p_K^t \times p_K^t)$ is the covariance of $p_K(N_\delta)$. (Here K and K^{**} are identified through their natural isomorphism.) Furthermore, note that $\delta_K\circ(p_K \times p_K) = \delta\circ(q_K \times q_K)$, where $q_K:V \to V$ is the orthogonal projection onto $p_K^{-1}(0)^\perp$ wrt δ, (cf. Section 2.3), hence $\delta_K\circ(p_K \times p_K)$ does not depend on the representation of K in its equivalence class of quotients.

The following important fact often has appeared in the literature in other forms.

---

[1]In a multivariate normal distribution, conditional distributions are more easily expressed in terms of the precision δ than in terms of its inverse $\delta^{-1} \equiv \sigma$ (the covariance).

Proposition 3.1. For any pair of quotients L,M of V, the following three conditions are equivalent:

(i) Under the distribution N_δ, p_L and p_M are conditionally independent (CI) given $p_{L\wedge M}$.

(ii) $\delta_{LVM}\circ(p_{LVM}\times p_{LVM}) = \delta_L\circ(p_L\times p_L) + \delta_M\circ(p_M\times p_M) - \delta_{L\wedge M}\circ(p_{L\wedge M}\times p_{L\wedge M})$.

(iii) L and M are geometrically orthogonal (GO) wrt δ.

Proof: Let $(e_i|i\in I)$ be a basis for V adapted to the distributive lattice $\mathcal{K} = \{L\wedge M, L, M, LVM, V\}$ (cf. Remark 2.2 and Figure 3.6) and let $R = I_L$, $C = I_M$. If the index set I is partioned as the disjoint union $I = (R\cup C)\,\dot{\cup}\,(I\setminus(R\cup C))$ then the matrix for $\delta_{LVM}\circ(p_{LVM}\times p_{LVM})$ wrt $(e_i|i\in I)$ takes the form

$$\begin{bmatrix}\Delta_{LVM} & 0\\ 0 & 0\end{bmatrix}, \text{ where } \Delta_{LVM} \equiv \begin{bmatrix}\Delta_{11} & \Delta_{12} & \Delta_{13}\\ \Delta_{21} & \Delta_{22} & \Delta_{23}\\ \Delta_{31} & \Delta_{32} & \Delta_{33}\end{bmatrix}$$

is the $(R\cup C)\times(R\cup C)$ matrix for δ_{LVM} wrt $(e_i|i\in R\cup C)$, and where subscripts 1, 2, and 3 correspond to the partitioning $R\cup C = (R\cap C)\dot{\cup}(R\setminus C)\dot{\cup}(C\setminus R)$. Similary, the matrix for $\delta_L\circ(p_L\times p_L)$ is

$$\begin{bmatrix}\Delta_L & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}, \text{ where } \Delta_L := \begin{bmatrix}\Delta_{11} & \Delta_{12}\\ \Delta_{21} & \Delta_{22}\end{bmatrix} - \begin{bmatrix}\Delta_{13}\\ \Delta_{23}\end{bmatrix}\Delta_{33}^{-1}\begin{bmatrix}\Delta_{31} & \Delta_{32}\end{bmatrix}$$

is the R×R matrix for δ_L wrt $(e_i|i\in R)$; the matrix for $\delta_M\circ(p_M\times p_M)$ is

$$\begin{bmatrix} \Lambda_M^{11} & 0 & \Lambda_M^{13} & 0 \\ 0 & 0 & 0 & 0 \\ \Lambda_M^{31} & 0 & \Lambda_M^{33} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \text{ where } \Lambda_M := \begin{bmatrix} \Lambda_M^{11} & \Lambda_M^{13} \\ \Lambda_M^{31} & \Lambda_M^{33} \end{bmatrix} = \begin{bmatrix} \Lambda_{11} & \Lambda_{13} \\ \Lambda_{13} & \Lambda_{33} \end{bmatrix} - \begin{bmatrix} \Lambda_{12} \\ \Lambda_{32} \end{bmatrix} \Lambda_{22}^{-1} \begin{bmatrix} \Lambda_{21} & \Lambda_{23} \end{bmatrix}$$

is the C×C matrix for δ_M wrt $(e_i | i \in C)$; and the matrix for $\delta_{L \wedge M} \circ (p_{L\wedge M} \times p_{L\wedge M})$ is

$$\begin{bmatrix} \Lambda_{L\wedge M} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \text{ where } \Lambda_{L\wedge M} := \Lambda_{11} - (\Lambda_{12}\ \Lambda_{13}) \begin{bmatrix} \Lambda_{22} & \Lambda_{23} \\ \Lambda_{32} & \Lambda_{33} \end{bmatrix}^{-1} \begin{bmatrix} \Lambda_{21} \\ \Lambda_{31} \end{bmatrix}$$

is the $(R\cap C)\times(R\cap C)$ matrix for $\delta_{L\wedge M}$ wrt $(e_i | i \in R\cap C)$.

(i) <=> (ii): From the relations $p_L = p_{L,LVM} \circ p_{LVM}$, $p_M = p_{M,LVM} \circ p_{LVM}$, and $p_{L\wedge M} = p_{L\wedge M,LVM} \circ p_{LVM}$ it is clear that (i) is equivalent to the condition that under the marginal distribution $p_{LVM}(N_\delta)$ on LVM, $p_{L,LVM}$ and $p_{M,LVM}$ are CI given $p_{L\wedge M,LVM}$. Since $p_{LVM}(N_\delta)$ has precision δ_{LVM}, (i) is thus equivalent to Λ_{23} $(= \Lambda_{32}^t) = 0$. It is seen from the above matrix representations, however, that $\Lambda_{23} = 0$ <=> (ii) is valid.

(ii) <=> (iii): Both formulas are purely algebraic and so is the proof of their equivalence. Clearly (ii) holds if and only if

$$\delta \circ (q_{LVM} \times q_{LVM}) = \delta \circ (q_L \times q_L) + \delta \circ (q_M \times q_M) - \delta \circ (q_{L\wedge M} \times q_{L\wedge M}).$$

Since q_{LVM} and $q_{L\wedge M}$ are the orthogonal projections onto $(p_L^{-1}(0) \cap p_M^{-1}(0))^\perp$ and $(p_L^{-1}(0) + p_M^{-1}(0))^\perp$, respectively, this relation is equivalent to

$$\delta\circ(q_{L\vee M}\times q_{L\vee M}) =$$
$$\delta\circ((q_L-q_{L\wedge M})\times(q_L-q_{L\wedge M})) + \delta\circ((q_M-q_{L\wedge M})\times(q_M-q_{L\wedge M})) + \delta\circ(q_{L\wedge M}\times q_{L\wedge M}),$$

which holds if and only if $p_L^{-1}(0)^\perp$ and $p_M^{-1}(0)^\perp$ are GO, i.e., L and M are GO wrt δ. □

For every sublattice $\mathcal{K} \subseteq \mathcal{K}(V)$ such that $V \in \mathcal{K}$, let $\mathbf{P}^{\mathcal{K}}(V)$ denote the set of all positive definite forms δ on V such that L and M are GO wrt δ for all $L,M \in \mathcal{K}$. Also, let $\mathbf{P}_{\mathcal{K}}(V) = [\mathbf{P}^{\mathcal{K}}(V)]^{-1}$, the corresponding set of positive definite forms on the dual space V^*. If $\delta \in \mathbf{P}^{\mathcal{K}}(V)$ represents the precision of a multivariate normal distribution on V, then $\sigma \equiv \delta^{-1} \in \mathbf{P}_{\mathcal{K}}(V)$ is its covariance.

It follows from Theorem 2.2 that $\mathbf{P}^{\mathcal{K}}(V) \neq \emptyset$ if and only if $\mathcal{K}$ is distributive. Therefore, in the remainder of this paper we shall consider only lattices $\mathcal{K} \subseteq \mathcal{K}(V)$ that are distributive (hence finite).

Remark 3.1. If $V = \mathbb{R}^I$ where I is a finite set, then we identify $\mathbf{P}(\mathbb{R}^I)$ with $\mathbf{P}(I)$, the set of all positive definite I×I matrices, in the usual way. If $\mathcal{R}$ is a ring of subsets of I such that $I \in \mathcal{R}$ (cf. Example 2.1) then, under this identification, $\mathbf{P}^{\mathcal{K}(\mathcal{R})}(\mathbb{R}^I)$ becomes a subset $\mathbf{P}^{\mathcal{R}}(I)$ of $\mathbf{P}(I)$. For the purpose of Section 3.3, we also define $\mathbf{P}_{\mathcal{R}}(I) := [\mathbf{P}^{\mathcal{R}}(I)]^{-1}$.□

Remark 3.2. For $R \in \mathcal{R}$ define $R' = I\setminus R$ and set $\mathcal{R}' = \{R' \mid R\in\mathcal{R}\}$. Then $\mathcal{R}'$ is also a ring of subsets of I and $\mathbf{P}^{\mathcal{R}'}(I) = \mathbf{P}_{\mathcal{R}}(I)$, $\mathbf{P}_{\mathcal{R}'}(I) = \mathbf{P}^{\mathcal{R}}(I)$. □

By Proposition 3.1, $P^{\mathcal{K}}(V)$ is the set of all precisions δ for which, under N_δ, p_L and p_M are CI given $p_{L\wedge M}$ for every pair $L,M \in \mathcal{K}$. The following theorem characterizes $P^{\mathcal{K}}(V)$:

Theorem 3.1. Let δ be a positive definite form on V. Then the following three conditions are equivalent:

(i) $\delta \in P^{\mathcal{K}}(V)$.

(ii) $\delta(x) = \Sigma(\delta_K(p_K(x))-\delta_{J(K)}(p_{J(K)}(x))\,|K\in J(\mathcal{K}))$, $x \in V$.

(iii) $\delta_L(p_L(x)) = \Sigma(\delta_K(p_K(x))-\delta_{J(K)}(p_{J(K)}(x))\,|K\in J(\mathcal{K}),K\leq L)$, $x \in V$, $L \in \mathcal{K}$.

Proof: That (iii) => (ii) is trivial. To show that (ii) => (iii), first note that (ii) can be rewritten as

$$(3.1)\qquad \begin{aligned}\delta(x) &= \Sigma(\delta(q_K(x))-\delta(q_{J(K)}(x))\,|K\in J(\mathcal{K}))\\ &= \Sigma(\delta(q_K(x)-q_{J(K)}(x))\,|K\in J(\mathcal{K})),\end{aligned}$$

where the second equality follows from the relation $p_{J(K)}^{-1}(0)^\perp \subseteq p_K^{-1}(0)^\perp$, when $K \in J(\mathcal{K})$. By Theorem 2.1 of [A] (1987) with $\mathcal{L} = \{p_K^{-1}(0)^\perp|K\in\mathcal{K}\}$ and $V_L = p_K^{-1}(0)^\perp\cap\, p_{J(K)}^{-1}(0)$ for $L \equiv p_K^{-1}(0)^\perp \in J(\mathcal{L})$,

$$(3.2)\qquad V = \oplus(p_K^{-1}(0)^\perp\cap\, p_{J(K)}^{-1}(0)\,|K\in J(\mathcal{K})).$$

Since $q_K-q_{J(K)}$ is the orthogonal projection onto $p_K^{-1}(0)^\perp\cap\, p_{J(K)}^{-1}(0)$ wrt δ, $K \in J(\mathcal{K})$, it follows from (3.1) that the direct sum is orthogonal wrt δ. Thus for $L \in \mathcal{K}$ we have that

$$(3.3)\qquad p_L^{-1}(0)^\perp = \oplus(p_K^{-1}(0)^\perp \cap p_{J(K)}^{-1}(0)\,|K\in J(\mathscr{K}),\ K\le L)$$

(cf. (2.5) of [A] (1987)) and this direct sum is again orthorgonal. It then follows that

$$q_L = \Sigma(q_K - q_{J(K)}\,|K\in J(\mathscr{K}), K\le L),$$

hence

$$\begin{aligned}\delta_L(p_L(x)) &= \delta(q_L(x))\\ &= \delta(\Sigma((q_K-q_{J(K)})(x)\,|K\in J(\mathscr{K}), K\le L)\\ &= \Sigma(\delta(p_K(x))-\delta(p_{J(K)}(x))\,|K\in J(\mathscr{K}), K\le L),\end{aligned}$$

so (ii) => (iii).

It was seen in the above argument that when (ii) holds, the direct sums (3.2) and (3.3) are orthogonal wrt δ, hence $\mathscr{L}$ consists of GO subspaces, i.e., $\delta \in \mathbf{P}^{\mathscr{K}}(V)$. Conversely, if $\delta \in \mathbf{P}^{\mathscr{K}}(V)$ then the direct sum (3.2) is orthogonal wrt δ (cf. Theorem 2.1 of [A] (1987)), hence

$$\begin{aligned}\delta(x) &= \Sigma(\delta(q_K(x)-q_{J(K)}(x))\,|K\in J(\mathscr{K}))\\ &= \Sigma(\delta(q_K(x))-\delta(q_{J(K)}(x))\,|K\in J(\mathscr{K}))\\ &= \Sigma(\delta_K(p_K(x))-\delta_{J(K)}(p_{J(K)}(x))\,|K\in J(\mathscr{K})).\end{aligned}$$ □

Remark 3.3. Condition (iii) in Theorem 3.1 is equivalent to

$$(\mathrm{iii})'\qquad \delta_K(p_K(x))-\delta_{J(K)}(p_{J(K)}(x)) = \Sigma(\mu(K,L)\delta_L(p_L(x))\,|L\in J(\mathscr{K})),$$

$K\in J(\mathscr{K})$, where μ is the Möbius function for $J(\mathscr{K})$. (cf. Lemma 1.1 of [A] (1987).) Condition (ii) of Proposition 3.1 is a special case of (iii)'. □

3.2. The $\mathcal{K}$-parametrization of $P^{\mathcal{K}}(V)$ and $P_{\mathcal{K}}(V)$.

Let $\mathbf{P}(n)$ be the cone of all $n\times n$ positive definite real matrices and let $\mathbf{M}(n\times m)$ be the vector space of all $n\times m$ real matrices. For any partitioning $n = n_1+n_2$ it is well known that $\mathbf{P}(n)$ can be parametrized by the product $\mathbf{P}(n_1)\times\mathbf{M}(n_1\times n_2)\times\mathbf{P}(n_2)$ under the 1-1 correspondence

$$(3.4) \qquad \begin{array}{ccc} \mathbf{P}(n) & \longleftrightarrow & \mathbf{P}(n_1)\times\mathbf{M}(n_1\times n_2)\times\mathbf{P}(n_2) \\ \Sigma = \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix} & \to & (\Sigma_{11},\ \Sigma_{21}\Sigma_{11}^{-1},\ \Sigma_{22\cdot 1}) \\ \begin{bmatrix}\Omega & \Omega R^t\\ R\Omega & \Lambda+R\Omega R^t\end{bmatrix} & \leftarrow & (\Omega,R,\Lambda). \end{array}$$

It will be seen from Theorem 3.2 and (3.13) that (3.4) is an example (in matrix formulation) of the $\mathcal{K}$-<u>parametrization</u> of $P_{\mathcal{K}}(V)$ for a general distributive lattice $\mathcal{K} \subseteq \mathcal{K}(V)$; in this example $(\Sigma_{11},\ \Sigma_{21}\Sigma_{11}^{-1},\ \Sigma_{22\cdot 1})$ are the $\mathcal{K}$-<u>parameters</u> of Σ. The $\mathcal{K}$-parametrization is central for the statistical analysis of the normal model with parameter space $P_{\mathcal{K}}(V)$. (cf. Sections 5 and 6).

For $K \in J(\mathcal{K})$, let $r_K:K \to p^{-1}_{J(K),K}(0)$ be a <u>projection</u>, i.e., a surjective linear mapping with the property $r_K(r_K(x)) = r_K(x)$, $x \in K$. If also $\delta \in P^{\mathcal{K}}(V)$, let $q^{\delta}_{J(K),K}:K \to K$ denote the orthogonal projection of K onto $p^{-1}_{J(K),K}(0)^{\perp}$ wrt δ_K, let $r^{\delta}_K:K \to p^{-1}_{J(K),K}(0)$ denote the mapping defined by $r^{\delta}_K(x) = x - q^{\delta}_{J(K),K}(x)$, $x \in K$, and let γ^{δ}_K denote the restriction of δ_K to $p^{-1}_{J(K),K}(0)$. Furthermore, let $\mathbf{PROJ}(K,p^{-1}_{J(K),K}(0))$ and $\mathbf{P}(p^{-1}_{J(K),K}(0))$ denote the set of all projections of K onto $p^{-1}_{J(K),K}(0)$ and the set of all positive definite forms on $p^{-1}_{J(K),K}(0)$, respectively. Finally, let $id_K:K \to K$ denote the identity mapping.

<u>Theorem 3.2.</u> ($\mathcal{H}$-parametrizations of $\mathbf{P}^{\mathcal{H}}(V)$ and $\mathbf{P}_{\mathcal{H}}(V)$). The mappings

$$\text{(3.5)} \qquad \mathbf{P}^{\mathcal{H}}(V) \longleftrightarrow X(\mathbf{PROJ}(K, p^{-1}_{J(K),K}(0)) \times \mathbf{P}(p^{-1}_{J(K),K}(0)) \,|\, K \in J(\mathcal{H}))$$

$$\text{(3.5a)} \qquad \delta \to ((r^{\delta}_K, \gamma^{\delta}_K) \,|\, K \in J(\mathcal{H}))$$

$$\text{(3.5b)} \qquad (x \to \Sigma(\gamma_K(r_K(p_K(x)))) \,|\, K \in J(\mathcal{H})) \leftarrow ((r_K, \gamma_K) \,|\, K \in J(\mathcal{H}))$$

define a 1-1 correspondence, called the $\mathcal{H}$-<u>parametrization</u> of $\mathbf{P}^{\mathcal{H}}(V)$. The $\mathcal{H}$-<u>parametrization</u> of $\mathbf{P}_{\mathcal{H}}(V)$ is determined by (3.5) through the inverse mapping $\delta \to \delta^{-1} \equiv \sigma$. (See Section 3.3 for the $\mathcal{H}$-parametrization of $\mathbf{P}_{\mathcal{H}}(V)$ in matrix formulation.).

<u>Proof:</u> Clearly the image of the mapping (3.5a) lies in the product set in (3.5). To show that the image of (3.5b) lies in $\mathbf{P}^{\mathcal{H}}(V)$, define the positive semidefinite form δ on V by

$$\text{(3.6)} \qquad \delta(x) = \Sigma(\gamma_K(r_K(p_K(x)))) \,|\, K \in J(\mathcal{H})).$$

It follows from (2.1) that $\Sigma(\gamma_K | K \in J(\mathcal{H})) \circ (\varphi_V \times \varphi_V) = \delta$, where $\Sigma(\gamma_K | K \in J(\mathcal{H}))$ is the positive definite form on $X(p^{-1}_{J(K),K}(0) \,|\, K \in J(\mathcal{H}))$ determined by

$$\Sigma(\gamma_K | K \in J(\mathcal{H}))((z_K | K \in J(\mathcal{H}))) = \Sigma(\gamma_K(z_K) \,|\, K \in J(\mathcal{H})).$$

By Theorem 2.1, δ is in fact positive definite, while for $L \in \mathcal{H}$

$$\delta_L = \Sigma(\gamma_K | K \in J(\mathcal{H}), K \leq L) \circ (\varphi_L \times \varphi_L),$$

by Remark 2.1. Thus, for $L \in J(\mathcal{H})$ and $x \in V$,

$$\begin{aligned}\delta_L(p_L(x)) &= \Sigma(\gamma_K(r_K(p_{KL}(p_L(x)))\,|K\in J(\mathcal{H}),K\leq L)\\ &= \Sigma(\gamma_K(r_K(p_K(x))\,|K\in J(\mathcal{H}),K\leq L)\end{aligned}$$

and

$$\delta_{J(L)}(p_{J(L)}(x)) = \Sigma(\gamma_K(r_K(p_K(x))\,|K\in J(\mathcal{H}),K\leq J(L)).$$

Since $\{K\in J(\mathcal{H})\,|K\leq L\} = \{K\in J(\mathcal{H})\,|K\lneq J(L)\}\dot{\cup}\{L\}$, it follows that

$$\delta_L(p_L(x)) - \delta_{J(L)}(p_{J(L)}(x)) = \gamma_L(r_L(p_L(x)). \tag{3.7}$$

Thus from Theorem 3.1(ii) it follows that $\delta \in \mathbf{P}^{\mathcal{H}}(V)$ as claimed.

Next we shall show that the composition of the mappings (3.5b) and (3.5a) is the identity on the product set in (3.5). Since, for every $L \in J(\mathcal{H})$ and $x \in V$,

$$\begin{aligned}\delta_L(p_L(x)) - \delta_{J(L)}(p_{J(L)}(x)) &= \delta_L(p_L(x)) - \delta_{J(L)}(p_{J(L),L}(p_L(x)))\\ &= \delta_L((id_L-q^{\delta}_{J(L),L})(p_L(x)))\\ &= \gamma^{\delta}_L(r^{\delta}_L(p_L(x))),\end{aligned} \tag{3.8}$$

it follows from (3.7) that

$$\gamma^{\delta}_L(r^{\delta}_L(y)) = \gamma_L(r_L(y)) \tag{3.9}$$

for every $y \in L$. However, for every $y \in p^{-1}_{J(L),L}(0)$ we have that $r^{\delta}_L(y) = r_L(y) = y$, hence $\gamma^{\delta}_L = \gamma_L$ by (3.9). Also, (3.9) implies that the kernels of r^{δ}_L and r_L coincide, hence $r^{\delta}_L = r_L$.

Finally, it follows from Theorem 3.1(ii) and (3.8) that

$$(3.10) \qquad \delta(x) = \Sigma(\gamma_K^{\delta}(r_K^{\delta}(p_K(x)))\,|K\in J(\mathcal{K})), \quad \delta \in \mathbf{P}^{\mathcal{K}}(V), \ x \in V,$$

hence the composition of the mappings (3.5a) and (3.5b) is the identity on $\mathbf{P}^{\mathcal{K}}(V)$, so the relation (3.5) is 1-1. □

Remark 3.4. When $K = 0_{\mathcal{K}}$, then $p_{J(K),K}^{-1}(0) = K$, hence $\mathbf{PROJ}(K,p_{J(K),K}^{-1}(0)) = \mathbf{PROJ}(K,K) = \{id_K\}$, so this term may be excluded from the product on the right of (3.5). □

Remark 3.5. The formula (3.10) is basic for likelihood inference for the normal statistical model with parameter space $\mathbf{P}_{\mathcal{K}}(V)$ - cf. (3.14) and Section 5. □

3.3. The $\mathcal{K}$-parametrization of $\mathbf{P}_{\mathcal{K}}(V)$: matrix formulation.

If D and E are finite index sets then $\mathbf{M}(D\times E)$ denotes the vector space of all real $D\times E$ matrices and $\mathbf{P}(D)$ the set of real positive definite $D\times D$ matrices. If $(e_i\,|i\in I)$ is a basis for V, then both $\mathbf{P}(V)$ and $\mathbf{P}(V^*)$ may be identified with $\mathbf{P}(I)$ in the usual way. If this basis is adapted to $\mathcal{K}$, then under this identification $\mathbf{P}^{\mathcal{K}}(V)$ becomes identified with $\mathbf{P}^{\mathfrak{R}(\mathcal{K})}(I) =: \mathbf{P}^{\mathcal{K}}(I)$ and $\mathbf{P}_{\mathcal{K}}(V)$ with $\mathbf{P}_{\mathfrak{R}(\mathcal{K})}(I) =: \mathbf{P}_{\mathcal{K}}(I)$ (cf. Remarks 2.2 and 3.1).

Since, for $K \in J(\mathcal{K})$, $(e_i\,|i\in I_K)$ and $(e_i\,|i\in D_K)$ are bases for K and $p_{J(K),K}^{-1}(0)$ respectively, it follows that $\mathbf{PROJ}(K,p_{J(K),K}^{-1}(0))$ may be identified with all $D_K\times I_K$ matrices of the form $(R_K, 1_{KK})$ partitioned according to (2.4), where $R_K \in \mathbf{M}(D_K \times I_{J(K)})$ and 1_{KK} is the $D_K \times D_K$ identity matrix, hence may be identified with $\mathbf{M}(D_K \times I_{J(K)})$. Also $\mathbf{P}(p^{-1}_{J(K),K}(0))$ may be identified with $\mathbf{P}(D_K)$.

For every $\Sigma \in \mathbf{P}(I)$ and $L \in \mathcal{K}$, denote the $I_L \times I_L$ submatrix of Σ by Σ_L and let Σ_L^{-1} denote $(\Sigma_L)^{-1}$. Also, for $K \in J(\mathcal{K})$, partition Σ_K according to (2.4) as follows[1]:

$$\Sigma_K = \begin{bmatrix} \Sigma_{\langle K\rangle} & \Sigma_{\langle K} \\ \Sigma_{K\rangle} & \Sigma_{KK} \end{bmatrix} \tag{3.11}$$

where $\Sigma_{\langle K\rangle} := \Sigma_{J(K)}$ is $I_{J(K)} \times I_{J(K)}$, $\Sigma_{K\rangle}$ is $D_K \times I_{J(K)}$, Σ_{KK} is $D_K \times D_K$, and $\Sigma_{\langle K} = \Sigma^t_{K\rangle}$. Furthermore, define

$$\Sigma_{K\bullet} = \Sigma_{KK} - \Sigma_{K\rangle}\Sigma^{-1}_{\langle K\rangle}\Sigma_{\langle K} \tag{3.12}$$

and let $\Sigma^{-1}_{K\bullet}$ denote $(\Sigma_{K\bullet})^{-1}$.

Under the identifications in the first paragraph, each $\Sigma \in \mathbf{P}_{\mathcal{K}}(I)$ corresponds uniquely to some $\delta \in \mathbf{P}^{\mathcal{K}}(V)$. The matrix wrt $(e_i \mid i \in I_K)$ for $(\delta_K)^{-1}$ is simply Σ_K, hence the matrix wrt $(e_i \mid i \in I_K)$ and $(e_i \mid i \in D_K)$ for r^δ_K is $(\Sigma_{K\rangle}\Sigma^{-1}_{\langle K\rangle}, 1_{KK})$. Since the matrix wrt $(e_i \mid i \in D_K)$ for $(\gamma^\delta_K)^{-1}$ is $\Sigma_{K\bullet}$, the $\mathcal{K}$-parametrization of $\mathbf{P}_{\mathcal{K}}(V)$ has the following matrix formulation:

$$\begin{aligned} \mathbf{P}_{\mathcal{K}}(I) &\longleftrightarrow \times(\mathbf{M}(D_K \times I_{J(K)}) \times \mathbf{P}(D_K) \mid K \in J(\mathcal{K})) \\ \Sigma &\to ((\Sigma_{K\rangle}\Sigma^{-1}_{\langle K\rangle}, \Sigma_{K\bullet}) \mid K \in J(\mathcal{K})). \end{aligned} \tag{3.13}$$

[1] If $K = 0_{\mathcal{K}}$, recall that $J(0_{\mathcal{K}}) = \{0\}$ and $I_{J(K)} = \emptyset$. In this case, $\Sigma_K = \Sigma_{KK} = \Sigma_{K\bullet}$.

The matrices $(\Sigma_{K>}\Sigma_{<K>}^{-1},\Sigma_{K\cdot})$, $K \in J(\mathcal{K})$, are called the $\mathcal{K}$-parameters of Σ.

The formula (3.10) now assumes the form

$$\text{(3.14)}\qquad \mathrm{tr}(\Sigma^{-1}xx^t) = \Sigma((\mathrm{tr}(\Sigma_{K\cdot}^{-1}(x_{[K]}-\Sigma_{K>}\Sigma_{<K>}^{-1}x_{<K>})(x_{[K]}-\Sigma_{K>}\Sigma_{<K>}^{-1}x_{<K>})^t)|K\in J(\mathcal{K}))$$

where, for $L \in \mathcal{K}$, x_L denotes the I_L-subcolumn of the column vector $x \in \mathbb{R}^I$ and, for $K \in J(\mathcal{K})$, x_K is partitioned according to (2.4) as

$$\text{(3.15)}\qquad x_K = \begin{bmatrix} x_{<K>} \\ x_{[K]} \end{bmatrix},$$

where $x_{<K>} := x_{J(K)}$. Formula (3.14) is used to obtain the maximum likelihood estimate of Σ under the normal model studied in Section 5.

Formula (3.14) provides a way to reconstruct Σ from its $\mathcal{K}$-parameters. If the $\mathcal{K}$-parameters are denoted by

$$\text{(3.16)}\qquad ((R_K,\Lambda_K)|K\in J(\mathcal{K})) \in X(\mathbf{M}(D_K\times I_{J(K)})\times\mathbf{P}(D_K)|K\in J(\mathcal{K}))$$

(compare to (3.13)), it follows directly from (3.14) that

$$\text{(3.17)}\qquad \Sigma = [\Sigma(\tilde{\Lambda}_K|K\in J(\mathcal{K}))]^{-1},$$

where $\tilde{\Lambda}_K$ is the $I\times I$ matrix whose $I_K\times I_K$ submatrix is

$$\begin{bmatrix} R_K^t\Lambda_K^{-1}R_K & \Lambda_K^{-1}R_K \\ R_K^t\Lambda_K^{-1} & \Lambda_K^{-1} \end{bmatrix}$$

and whose remaining entries are zeroes. Formulas (3.13) and (3.17) are the analogues of (3.5a) and (3.5b), respectively, for the covariance matrix Σ.

In general it is not a simple task to determine Σ directly from (3.17) by matrix inversion. In the following extended remark we present a step-wise algorithm for reconstructing Σ from its $\mathcal{K}$-parameters without matrix inversion.

Remark 3.6. Clearly the $\mathcal{K}$-parameters of Σ are simple functions of Σ. The process of reconstructing Σ from its $\mathcal{K}$-parameters is, in general, more complex ((3.4) is a very simple case). It is important to describe this process since, as will be seen in Section 5, the maximum likelihood estimator $\hat{\Sigma}$ is obtained by first estimating the $\mathcal{K}$-parameters of Σ, then using them to obtain $\hat{\Sigma}$.

We now present a step-wise algorithm for this reconstruction process. Let $n := |J(\mathcal{K})|$ and $0_{\mathcal{K}} \equiv K_1,\cdots,K_n$ be a never-decreasing listing of the members of the poset $J(\mathcal{K})$, i.e., $K_i \not> K_j$ whenever $i < j$; usually we shall abbreviate K_k to k. Partition Σ according to the ordered decomposition

$$I = D_1 \dot{\cup} D_2 \dot{\cup} \cdots \dot{\cup} D_n, \tag{3.18}$$

where $D_k \equiv D_{K_k}$. To reconstruct Σ from

(3.19)
$$(\Lambda_1, (R_2, \Lambda_2), \cdots, (R_n, \Lambda_n)) \in$$
$$\mathbf{P}(D_1)\times\mathbf{M}(D_2\times I_{J(2)})\times\mathbf{P}(D_2)\times\cdots\times\mathbf{M}(D_n\times I_{J(n)})\times\mathbf{P}(D_n)$$

where $J(k) \equiv J(K_k)$, one proceeds as follows:

Step 1:
$$\Sigma_1 = \Lambda_1.$$

Step 2:
$$\Sigma_{2\rangle} = R_2\Sigma_1,$$
$$\Sigma_{22} = \Lambda_2 + R_2\Sigma_{\langle 2}.$$

At this point, the submatrix Σ_{1V2} (= Σ_2, here) is completely determined (recall that 1V2 abbreviates $K_1 V K_2$). By (3.20), $J(3) \leq K_1 V K_2$ (= K_2, here), hence $I_{J(3)} \subseteq I_{1V2}$ (= I_2, here), so $\Sigma_{\langle 3\rangle}$ is a submatrix of Σ_{1V2} and the next step may be carried out:

Step 3a:
$$\Sigma_{3\rangle} = R_3\Sigma_{\langle 3\rangle},$$
$$\Sigma_{33} = \Lambda_3 + R_3\Sigma_{\langle 3}.$$

It is important to note that after Steps 1, 2, and 3a, the three submatrices Σ_1, Σ_2, Σ_3 are now determined but the complete submatrix Σ_{1V2V3} may not yet be fully determined. The remaining $D_3\times(I_{1V2V3}\setminus I_3)$ - submatrix $\Sigma_{3\}}$ of Σ_{1V2V3} is determined from Σ_{1V2} by means of the GO (≡ pairwise CI) requirements imposed by $\mathcal{K}$ (cf. (3.22)):

Step 3b:
$$\Sigma_{3\rangle\!\rangle} = R_3\Sigma_{\langle 3\rangle\!\rangle}$$
$$(= \Sigma_{3\rangle}\Sigma_{\langle 3\rangle}^{-1}\Sigma_{\langle 3\rangle\!\rangle}),$$

where $\Sigma_{\langle 3\rangle\!\rangle}$ is the $I_{J(3)}\times(I_{1V2V3}\setminus I_3)$ - submatrix which, by (3.20) and (3.21), is in fact a submatrix of Σ_{1V2}.

After k-1 such steps, the submatrix $\Sigma_{1V\cdots V(k-1)}$ is fully determined and in turn may be used to obtain $\Sigma_{1V\cdots Vk}$. First note that the never-decreasing nature of $K_1,\cdots,K_n$ implies that

$$I_{1V\cdots Vk} = \dot{\cup}(D_j\,|\,j=1,\cdots,k)$$
$$I_k = \dot{\cup}(D_j\,|\,j=1,\cdots,k;\ K_j\leq K_k).$$

From these relations and (2.4) it may be deduced that

$$I_{J(k)} = I_k \cap I_{1V\cdots V(k-1)} \subseteq I_{1V\cdots V(k-1)} \tag{3.20}$$
$$I_{1V\cdots Vk}\setminus I_k = I_{1V\cdots V(k-1)}\setminus I_{J(k)} \subseteq I_{1V\cdots V(k-1)}. \tag{3.21}$$

Thus, if we denote the $D_k\times(I_{1V\cdots Vk}\setminus I_k)$ submatrix of Σ by $\Sigma_{k\rangle\!\rangle}$ and the $I_{J(k)}\times(I_{1V\cdots Vk}\setminus I_k)$ submatrix by $\Sigma_{\langle k\rangle\!\rangle}$, it follows from (3.20) and (3.21) that both $\Sigma_{\langle k\rangle}$ and $\Sigma_{\langle k\rangle\!\rangle}$ are in fact submatrices of $\Sigma_{1V\cdots V(k-1)}$, so the next step may be carried out:

Step k:
$$\Sigma_{k\rangle} = R_k\Sigma_{\langle k\rangle},$$
$$\Sigma_{kk} = \Lambda_k + R_k\Sigma_{\langle k},$$
$$\Sigma_{k\rangle\!\rangle} = R_k\Sigma_{\langle k\rangle\!\rangle} \tag{3.22}$$
$$(= \Sigma_{k\rangle}\Sigma_{\langle k\rangle}^{-1}\Sigma_{\langle k\rangle\!\rangle}).$$

The relation (3.22) is seen as follows. Since K_k (=: L) and $K_1 \vee \cdots \vee K_{k-1}$ (=: M) are GO wrt Σ^{-1}, it follows from Proposition 3.1(ii), (3.20), and (3.21) that the $D_k \times (I_{1 \vee \cdots \vee k} \setminus I_k)$ - submatrix of $(\Sigma_{1 \vee \cdots \vee k})^{-1}$ is a zero matrix, which is equivalent to (3.22).

The submatrix $\Sigma_{1 \vee \cdots \vee k}$ is fully determined after Step k; after n steps, $\Sigma_{1 \vee \cdots \vee n} = \Sigma$ is fully determined.

[In carrying out this algorithm one must use the convention that if $C \neq \emptyset$ and $D \neq \emptyset$, then the product of a $C \times \emptyset$ matrix with an $\emptyset \times D$ matrix is the $C \times D$ zero matrix.] □

3.4. Examples.

We now present a series of Examples to illustrate the following basic concepts: (i) the distributive lattice $\mathcal{K}$ of quotient spaces of the observation space V; (ii) identification of V with $\mathbb{R}^I$ by means of a $\mathcal{K}$-adapted basis for V determined by the poset $J(\mathcal{K})$ of join-irreducible elements of $\mathcal{K}$; (iii) the $\mathcal{K}$-parametrization of $\mathbf{P}_{\mathcal{K}}(I)$ and the specific form of (3.14); (iv) reconstruction of $\Sigma \in \mathbf{P}_{\mathcal{K}}(I)$ from its $\mathcal{K}$-parameters by means of the step-wise algorithm in Remark 3.6.

In each Example the lattice diagram of $\mathcal{K}$ appears in an accompanying Figure, in which the members of $J(\mathcal{K})$ are circled. In each Figure, the minimal element $0_{\mathcal{K}}$ appears at the left while the maximal element appears at the right. This apparently contradicts the convention in [A] (1987) where, in the lattice diagram for a lattice $\mathcal{L}$ of subspaces of V, the minimal element $0_{\mathcal{L}}$ appears at the right while V appears at the left. These two conventions are consistent, however, because of the anti-isomorphism (2.9) between $\mathcal{K}(V)$ and $\mathcal{L}(V)$ described in Remark 2.5.

These Examples also will be used in Section 4 to illustrate the notion of a $\mathcal{K}$-preserving mapping, in Section 5 to provide specific examples of normal statistical models determined by pairwise CI wrt $\mathcal{K}$, and in Section 6 where the problem of testing one model against another is treated.

Example 3.1. First consider the simple case where $\mathcal{K} = \{L,V\} \subseteq \mathcal{K}(V)$ (see Figure 3.1).

L o———o V

Figure 3.1

Since L and V trivially are GO wrt every $\delta \in P(V)$, $P^{\mathcal{K}}(V) = P(V)$ and $P_{\mathcal{K}}(V) = P(V^{*})$. In order to choose a basis $(e_i | i \in I)$ adapted to $\mathcal{K}$, note that $J(\mathcal{K}) = \{L,V\}$ and $J(L) = \{0\}$, $J(V) = L$. Thus the $\mathcal{K}$-parametrization (3.13) of $P_{\mathcal{K}}(I) = P(I)$ becomes

(3.23) $$P(I) \longleftrightarrow P(D_L) \times M(D_V \times I_L) \times P(D_V)$$

(3.23a) $$\begin{bmatrix} \Sigma_L & \Sigma_{<V} \\ \Sigma_{V>} & \Sigma_{VV} \end{bmatrix} \to (\Sigma_L, \Sigma_{V>}\Sigma_L^{-1}, \Sigma_{V\cdot})$$

where $\Sigma = \Sigma_V$ (cf. (3.11)), while the formula (3.14) becomes

(3.24) $$\mathrm{tr}(\Sigma^{-1}xx^t) = \mathrm{tr}(\Sigma_L^{-1}x_L x_L^t) + \mathrm{tr}(\Sigma_{V\cdot}^{-1}(x_{[V]} - \Sigma_{V>}\Sigma_L^{-1}x_L)(\cdots)^t).$$

The algorithm in Remark 3.6 for reconstructing Σ from its $\mathcal{K}$-parameters $(\Lambda_L, R_V, \Lambda_V)$ takes the form

<u>Step 1:</u> $$\Sigma_L = \Lambda_L$$

<u>Step 2:</u> $$\Sigma_{V>} = R_V \Sigma_L$$
$$\Sigma_{VV} = \Lambda_V + R_V \Sigma_{<V}$$

(compare to (3.4)). □

<u>Example 3.2.</u> If $\mathcal{K} = \{K_1, \cdots, K_n = V\} \subseteq \mathcal{K}(V)$ is an ascending chain, i.e., $K_1 < \cdots < K_n$, then a well-known generalization of the preceding example is obtained (see Figure 3.2).

$$K_1 \;\circ\!\!-\!\!\!-\!\!\circ\!\!-\!\!\!-\!\!\circ\!\!-\; \cdots \;-\!\circ\!\!-\!\!\!-\!\!\circ\; K_n = V$$
$$K_2 \quad K_3 \qquad\qquad K_{n-1}$$

Figure 3.2

Here again $\mathbf{P}^{\mathcal{K}}(V) = \mathbf{P}(V)$ and $\mathbf{P}_{\mathcal{K}}(V) = \mathbf{P}(V^*)$, but the $\mathcal{K}$-parametrizations are changed. To choose a $\mathcal{K}$-adapted basis, note that $J(\mathcal{K}) = \{K_1, \cdots, K_n\}$ and $J(K_1) = \{0\}$, $J(K_k) = K_{k-1}$, $k = 2, \cdots, n$. Then the $\mathcal{K}$-parametrization of $\mathbf{P}_{\mathcal{K}}(I) = \mathbf{P}(I)$ becomes

(3.25) $$\mathbf{P}(I) \longleftrightarrow \mathbf{P}(D_1) \times \mathbf{M}(D_2 \times I_1) \times \mathbf{P}(D_2) \times \cdots \times \mathbf{M}(D_n \times I_{n-1}) \times \mathbf{P}(D_n)$$
$$\Sigma \to (\Sigma_1,\ \Sigma_{2>}\Sigma_1^{-1},\ \Sigma_{2\cdot},\ \cdots,\ \Sigma_{n>}\Sigma_{n-1}^{-1},\ \Sigma_{n\cdot})$$

while (3.14) becomes

(3.26) $$\mathrm{tr}(\Sigma^{-1}xx^t) = \mathrm{tr}(\Sigma_1^{-1}x_1x_1^t) + \mathrm{tr}(\Sigma_{2\cdot}^{-1}(x_{[2]} - \Sigma_{2>}\Sigma_1^{-1}x_1)(\cdots)^t)$$
$$+ \cdots + \mathrm{tr}(\Sigma_{n\cdot}^{-1}(x_{[n]} - \Sigma_{n>}\Sigma_{n-1}^{-1}x_{n-1})(\cdots)^t),$$

where $K_1, K_2, \cdots, K_n$ are abbreviated as $1, 2, \cdots, n$ whenever they occur as subscripts. By Remark 3.6, Σ is reconstructed from the $\mathcal{K}$-parameters $(\Lambda_1, R_2, \Lambda_2, \cdots, R_n, \Lambda_n)$ as follows:

<u>Step 1:</u> $\quad \Sigma_1 = \Lambda_1$

<u>Step 2:</u> $\quad \Sigma_{2>} = R_2 \Sigma_1$

$$\Sigma_{22} = \Lambda_2 + R_2 \Sigma_{<2}$$

$\vdots$

<u>Step n:</u> $\quad \Sigma_{n>} = R_n \Sigma_{n-1}$

$$\Sigma_{nn} = \Lambda_n + R_n \Sigma_{<n}. \qquad \square$$

<u>Example 3.3.</u> Now consider the lattice $\mathcal{K} = \{\{0\} = L \wedge M, L, M, V = L \vee M\}$ (see Figure 3.3).

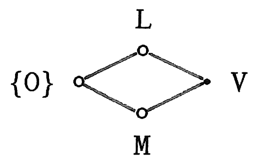

Figure 3.3

Here the GO ($\equiv$ pairwise CI) requirement imposed on Σ by $\mathcal{K}$ is nontrivial, so $\mathbf{P}_{\mathcal{K}}(V) \subset \mathbf{P}(V)$. In this case, note that $J(\mathcal{K}) = \{\{0\}, L, M\}$ and $J(\{0\}) = J(L) = J(M) = \{0\}$ and thereby choose a $\mathcal{K}$-adapted basis $(e_i \mid i \in I)$ for V. Then the $\mathcal{K}$-parametrization (3.13) of $\mathbf{P}_{\mathcal{K}}(I)$ takes the form

$$P_{\mathcal{K}}(I) \longleftrightarrow P(D_L) \times P(D_M) \qquad (3.27)$$
$$\Sigma \to (\Sigma_L,\ \Sigma_M),$$

while (3.14) becomes

$$\mathrm{tr}(\Sigma^{-1}xx^t) = \mathrm{tr}(\Sigma_L^{-1}x_Lx_L^t) + \mathrm{tr}(\Sigma_M^{-1}x_Mx_M^t). \qquad (3.28)$$

Since $\{0\}$, L, M is a never-decreasing listing of $J(\mathcal{K})$, Σ may be reconstructed from the nontrivial $\mathcal{K}$-parameters $(\Lambda_L,\ \Lambda_M)$ as follows:

Step 1: (vacuous)

Step 2: $\Sigma_L = \Lambda_L$

Step 3: $\Sigma_M = \Lambda_M$

$$\Sigma_{M\rangle} = 0.$$

Thus $P_{\mathcal{K}}(I)$ consists of all block-diagonal Σ of the form

$$\Sigma = \begin{bmatrix} \Sigma_L & 0 \\ 0 & \Sigma_M \end{bmatrix} \qquad (3.29)$$

where Σ is partitioned according to the decomposition

$$I = \emptyset \,\dot{\cup}\, D_L \,\dot{\cup}\, D_M. \qquad (3.30)$$

In this example (as in Examples 3.1 and 3.2) $P_{\mathcal{K}}(I) = P^{\mathcal{K}}(I)$ and both are linear, i.e., closed under nonnegative linear combinations. □

<u>Example 3.4.</u> If $\mathcal{K} = \{\{0\}=L\wedge M, L, M, L\vee M, V\}$ (see Figure 3.4)

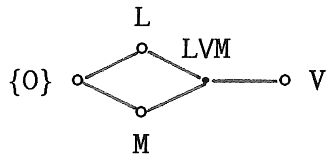

Figure 3.4

then again $P_{\mathcal{K}}(V) \subset P(V)$. Here note that $J(\mathcal{K}) = \{\{0\}, L, M, V\}$ and $J(\{0\}) = J(L) = J(M) = \{0\}$, $J(V) = L\vee M$, and thereby choose a $\mathcal{K}$-adapted basis. The $\mathcal{K}$-parametrization of $P_{\mathcal{K}}(I)$ assumes the form

(3.31)
$$\begin{gathered} P_{\mathcal{K}}(I) \longleftrightarrow P(D_L)\times P(D_M)\times M(D_V\times I_{L\vee M})\times P(D_V) \\ \Sigma \to (\Sigma_L,\ \Sigma_M,\ \Sigma_{V>}\Sigma_{L\vee M}^{-1},\ \Sigma_{V\cdot}) \end{gathered}$$

and (3.14) takes the form

$$tr(\Sigma^{-1}xx^t) = tr(\Sigma_L^{-1}x_Lx_L^t) + tr(\Sigma_M^{-1}x_Mx_M^t) + tr(\Sigma_{V\cdot}^{-1}(x_{[V]}-\Sigma_{V>}\Sigma_{L\vee M}^{-1}x_{L\vee M})(\cdots)^t)$$

Now $\{0\}, L, M, V$ is a never-decreasing listing of $J(\mathcal{K})$, so Σ may be reconstructed from the nontrivial $\mathcal{K}$-parameters $(\Lambda_L, \Lambda_M, R_V, \Lambda_V)$ as follows:

<u>Step 1,2,3:</u> Repeat Steps 1,2,3 in Example 3.3.

<u>Step 4:</u>
$$\begin{aligned} \Sigma_{V>} &= R_V\mathrm{Diag}(\Lambda_L, \Lambda_M) \\ \Sigma_{VV} &= \Lambda_V + R_V\Sigma_{<V}. \end{aligned}$$

Thus $P_{\mathcal{K}}(I)$ consists of all $\Sigma \in P(I)$ of the form

(3.32)
$$\Sigma = \left[\begin{array}{cc:c} \Sigma_L & 0 & \\ 0 & \Sigma_M & \Sigma_{<V} \\ \hdashline \multicolumn{2}{c:}{\Sigma_{V>}} & \Sigma_{VV} \end{array}\right]$$

where Σ is partioned according to the decomposition

(3.33)
$$I = \emptyset \,\dot{\cup}\, D_L \,\dot{\cup}\, D_M \,\dot{\cup}\, D_V.$$

The precision matrices $\Lambda = \Sigma^{-1} \in P^{\mathcal{K}}(I)$ are characterized by the condition that $\Sigma_{LVM}^{-1} = \mathrm{Diag}(\Sigma_L^{-1}, \Sigma_M^{-1})$. Thus, unlike the preceding example, here $P_{\mathcal{K}}(I)$ is linear while $P^{\mathcal{K}}(I)$ is not. □

Example 3.5. Suppose that $\mathcal{K} = \{L\wedge M, L, M, V=LVM\}$ where $L\wedge M > 0$ (see Figure 3.5).

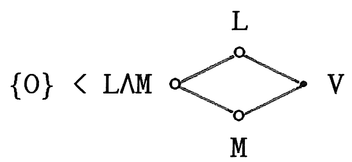

Figure 3.5

Note now that $J(\mathcal{K}) = \{L\wedge M, L, M\}$, and $J(L\wedge M) = \{0\}$, $J(L) = J(M) = L\wedge M$ and thereby choose a $\mathcal{K}$-adapted basis. The $\mathcal{K}$-parametrization of $P_{\mathcal{K}}(I)$ is given by

(3.34)
$$P_{\mathcal{K}}(I) \longleftrightarrow P(D_{L\wedge M}) \times M(D_L \times I_{L\wedge M}) \times P(D_L) \times M(D_M \times I_{L\wedge M}) \times P(D_M)$$
$$\Sigma \to (\Sigma_{L\wedge M},\ \Sigma_{L>}\Sigma_{L\wedge M}^{-1},\ \Sigma_{L\cdot},\ \Sigma_{M>}\Sigma_{L\wedge M}^{-1},\ \Sigma_{M\cdot}),$$

and (3.14) takes the form

$$\text{(3.35)} \qquad \begin{aligned} \operatorname{tr}(\Sigma^{-1}xx^t) &= \operatorname{tr}(\Sigma_{L\wedge M}^{-1}x_{L\wedge M}x_{L\wedge M}^t) \\ &+ \operatorname{tr}(\Sigma_{L\cdot}^{-1}(x_{[L]}-\Sigma_{L>}\Sigma_{L\wedge M}^{-1}x_{L\wedge M})(\cdots)^t) \\ &+ \operatorname{tr}(\Sigma_{M\cdot}^{-1}(x_{[M]}-\Sigma_{M>}\Sigma_{L\wedge M}^{-1}x_{L\wedge M})(\cdots)^t). \end{aligned}$$

Since $L\wedge M, L, M$ is a never-decreasing listing of $J(\mathcal{K})$, Σ may be reconstructed from the $\mathcal{K}$-parameters $(\Lambda_{L\wedge M}, R_L, \Lambda_L, R_M, \Lambda_M)$ as follows:

<u>Step 1:</u> $\Sigma_{L\wedge M} = \Lambda_{L\wedge M}$

<u>Step 2:</u> $\Sigma_{L>} = R_L\Sigma_{L\wedge M}$

$\Sigma_{LL} = \Lambda_L + R_L\Sigma_{<L}$

<u>Step 3:</u> $\Sigma_{M>} = R_M\Sigma_{L\wedge M}$

$\Sigma_{MM} = \Lambda_M + R_M\Sigma_{<M}$

(3.36) $\Sigma_{M\}} = R_M\Sigma_{<L}$

$= \Sigma_{M>}\Sigma_{L\wedge M}^{-1}\Sigma_{<L}$

(note that $\Sigma_{<M\}} = \Sigma_{<L}$). Thus, in this example $\mathbf{P}_{\mathcal{K}}(I)$ consists of all $\Sigma \in$ $\mathbf{P}(I)$ of the form

$$\text{(3.37)} \qquad \Sigma = \begin{bmatrix} \Sigma_{L\wedge M} & \Sigma_{<L} & \Sigma_{<M} \\ \Sigma_{L>} & \Sigma_{LL} & \Sigma_{\{M} \\ \Sigma_{M>} & \Sigma_{M\}} & \Sigma_{MM} \end{bmatrix}$$

such that $\Sigma_{M\}}$ satisfies (3.36) and where Σ is partitioned according to the decomposition

$$I = D_{L\wedge M} \dot{\cup}\, D_L \dot{\cup}\, D_M . \tag{3.38}$$

It is easily seen that $\mathbf{P}^{\mathcal{K}}(I)$ consists of all $\Lambda \in \mathbf{P}(I)$ having the simple form

$$\Lambda = \begin{bmatrix} \Lambda_{L\wedge M} & \Lambda_{<L} & \Lambda_{<M} \\ \Lambda_{L>} & \Lambda_{LL} & 0 \\ \Lambda_{M>} & 0 & \Lambda_{MM} \end{bmatrix} . \tag{3.39}$$

Thus, in this example $\mathbf{P}^{\mathcal{K}}(I)$ is linear while $\mathbf{P}_{\mathcal{K}}(I)$ is not. □

<u>Example 3.6.</u> Consider $\mathcal{K} = \{L\wedge M, L, M, L\vee M, V\}$ where $L\wedge M > \{0\}$ (see Figure 3.6).

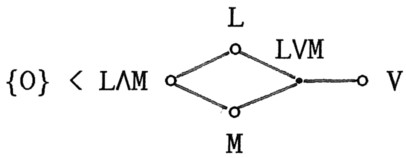

Figure 3.6

Choose a $\mathcal{K}$-adapted basis by noting that $J(\mathcal{K}) = \{L\wedge M, L, M, V\}$ and $J(L\wedge M) = \{0\}$, $J(L) = J(M) = L\wedge M$, $J(V) = L\vee M$. The $\mathcal{K}$-parametrization of $\mathbf{P}_{\mathcal{K}}(I)$ is given by

$$\mathbf{P}_{\mathcal{K}}(I) \longleftrightarrow \tag{3.40}$$

$$\mathbf{P}(D_{L\wedge M})\times\mathbf{M}(D_L\times I_{L\wedge M})\times\mathbf{P}(D_L)\times\mathbf{M}(D_M\times I_{L\wedge M})\times\mathbf{P}(D_M)\times\mathbf{M}(D_V\times I_{L\vee M})\times\mathbf{P}(D_V)$$

$$\Sigma \rightarrow$$

$$(\Sigma_{L\wedge M},\ \Sigma_{L>}\Sigma_{L\wedge M}^{-1},\ \Sigma_{L\cdot},\ \Sigma_{M>}\Sigma_{L\wedge M}^{-1},\ \Sigma_{M\cdot},\ \Sigma_{V>}\Sigma_{L\vee M}^{-1},\ \Sigma_{V\cdot})$$

while (3.14) becomes

$$(3.41)\qquad \begin{aligned} tr(\Sigma^{-1}xx^t) &= tr(\Sigma_{L\wedge M}^{-1}x_{L\wedge M}x_{L\wedge M}^t) \\ &+ tr(\Sigma_{L\cdot}^{-1}(x_{[L]}-\Sigma_{L>}\Sigma_{L\wedge M}^{-1}x_{L\wedge M})(\cdots)^t) \\ &+ tr(\Sigma_{M\cdot}^{-1}(x_{[M]}-\Sigma_{M>}\Sigma_{L\wedge M}^{-1}x_{L\wedge M})(\cdots)^t) \\ &+ tr(\Sigma_{V\cdot}^{-1}(x_{[V]}-\Sigma_{V>}\Sigma_{LVM}^{-1}x_{LVM})(\cdots)^t). \end{aligned}$$

Since $L\wedge M, L, M, V$ is a never-decreasing listing of $J(\mathscr{K})$, Σ can be reconstructed from $(\Lambda_{L\wedge M}, R_L, \Lambda_L, R_M, \Lambda_M, R_V, \Lambda_V)$ as follows:

Steps 1,2,3: Repeat Steps 1,2 and 3 in Example 2.5, obtaining Σ_{LVM}.

Step 4:
$$\Sigma_{V>} = R_V\Sigma_{LVM}$$
$$\Sigma_{VV} = \Lambda_V + R_V\Sigma_{<V}$$

Thus $P_{\mathscr{K}}(I)$ consists of all $\Sigma \in P(I)$ of the form

$$(3.42)\qquad \Sigma = \begin{bmatrix} \Sigma_{LVM} & \Sigma_{<V} \\ \Sigma_{V>} & \Sigma_{VV} \end{bmatrix}$$

partitioned according to $I = I_{LVM} \dot{\cup} D_V$, where Σ_{LVM} is given by (3.37), (3.36) and (3.38). The precision matrices $\Lambda \equiv \Sigma^{-1} \in P^{\mathscr{K}}(I)$ are characterized by the condition that Σ_{LVM}^{-1} have the form (3.39). Thus neither $P_{\mathscr{K}}(I)$ or $P^{\mathscr{K}}(I)$ are linear. □

<u>Example 3.7.</u> Let $\mathcal{K}$ be the lattice in Figure 3.7:

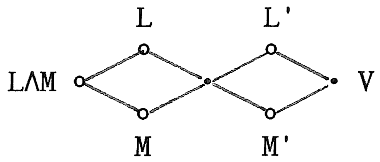

Figure 3.7.

Choose a $\mathcal{K}$-adapted basis by noting that $J(\mathcal{K}) = \{L\wedge M,\ L,\ M,\ L',\ M'\}$ and $J(L\wedge M) = \{0\}$, $J(L) = J(M) = L\wedge M$, $J(L') = J(M') = L\vee M = L' \wedge M'$. The $\mathcal{K}$-parametrization of $\mathbf{P}_{\mathcal{K}}(I)$ is given by

(3.43)
$$\mathbf{P}_{\mathcal{K}}(I) \longleftrightarrow$$
$$\mathbf{P}(D_{L\wedge M})\times\mathbf{M}(D_L\times I_{L\wedge M})\times\mathbf{P}(D_L)\times\mathbf{M}(D_M\times I_{L\wedge M})\times\mathbf{P}(D_M)$$
$$\times\mathbf{M}(D_{L'}\times I_{L\vee M})\times\mathbf{P}(D_{L'})\times\mathbf{M}(D_{M'}\times I_{L\vee M})\times\mathbf{P}(D_{M'})$$
$$\Sigma \to$$
$$(\Sigma_{L\wedge M},\ \Sigma_{L>}\Sigma_{L\wedge M}^{-1},\ \Sigma_{L\cdot},\ \Sigma_{M>}\Sigma_{L\wedge M}^{-1},\ \Sigma_{M\cdot},\ \Sigma_{L'>}\Sigma_{L\vee M}^{-1},\ \Sigma_{L'\cdot},\ \Sigma_{M'>}\Sigma_{L\vee M}^{-1},\ \Sigma_{M'\cdot})$$

from which the specific form of (3.14) is easily determined. The matrix Σ can be reconstructed from its $\mathcal{K}$-parameters $(\Lambda_{L\wedge M},\ R_L,\ \Lambda_L,\ R_M,\ \Lambda_M,\ R_{L'},\ \Lambda_{L'},\ R_{M'},\ \Lambda_{M'})$ as follows:

<u>Steps 1,2,3:</u> Repeat Steps 1,2,3 in Example 2.5, obtaining $\Sigma_{L\vee M} = (\Sigma_{L'\wedge M'})$.

<u>Steps 4,5:</u> Repeat Steps 2,3 in Example 2.5 with L,M replaced by L',M'.

Thus $\mathbf{P}_{\mathcal{K}}(I)$ consists of all $\Sigma \in \mathbf{P}(I)$ of the form (3.37) with L,M replaced by L',M', partitioned according to

$$(3.44) \qquad I = I_{L\vee M} \,\dot{\cup}\, D_{L'} \,\dot{\cup}\, D_{M'}$$

where furthermore, $\Sigma_{L'\wedge M'} = \Sigma_{L\vee M}$ is given by (3.37). The precision matrix $\Lambda = \Sigma^{-1}$ has the form (3.39) with L,M replaced by L',M' and satisfies the condition that $\Sigma^{-1}_{L\vee M}$ has the form (3.39). Here again neither $P_{\mathcal{K}}(I)$ nor $P^{\mathcal{K}}(I)$ are linear. □

<u>Example 3.8.</u> Let $\mathcal{K}$ be the lattice in Figure 3.8:

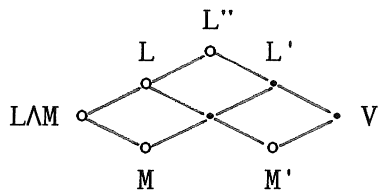

Figure 3.8.

Here $J(\mathcal{K}) = \{L\wedge M, L, M, L'', M'\}$ and $J(L\wedge M) = \{0\}$, $J(L) = J(M) = L\wedge M$, $J(L'') = L$, $J(M') = L\vee M = L'\wedge M'$. Thereby choose a $\mathcal{K}$-adapted basis and obtain the $\mathcal{K}$-parametrization of $P_{\mathcal{K}}(I)$ relative to such a basis:

$$(3.45) \qquad P_{\mathcal{K}}(I) \longleftrightarrow$$

$$P(D_{L\wedge M})\times\mathbf{M}(D_L\times I_{L\wedge M})\times P(D_L)\times\mathbf{M}(D_M\times I_{L\wedge M})\times P(D_M)$$
$$\times\mathbf{M}(D_{L''}\times I_L)\times P(D_{L''})\times\mathbf{M}(D_{M'}\times I_{L\vee M})\times P(D_{M'})$$
$$\Sigma \to$$
$$(\Sigma_{L\wedge M},\ \Sigma_{L>}\Sigma^{-1}_{L\wedge M},\ \Sigma_{L\cdot},\ \Sigma_{M>}\Sigma^{-1}_{L\wedge M},\ \Sigma_{M\cdot},\ \Sigma_{L''>}\Sigma^{-1}_{L},\ \Sigma_{L''\cdot},\ \Sigma_{M'>}\Sigma^{-1}_{L\vee M},\ \Sigma_{M'\cdot})$$

from which the specific form of (3.14) is determined. The matrix Σ can be reconstructed from its $\mathcal{K}$-parameters $(\Lambda_{L\wedge M}, R_L, \Lambda_L, R_M, \Lambda_M, R_{L''}, \Lambda_{L''}, R_{M'}, \Lambda_{M'})$ as follows:

<u>Steps 1,2,3:</u> Repeat Steps 1,2,3 in Example 2.5, obtaining $\Sigma_{L\vee M} = (\Sigma_{L'\wedge M'})$.

<u>Step 4:</u>

$$\Sigma_{L''\rangle} = R_{L''}\Sigma_{L}$$

$$\Sigma_{L''L''} = \Lambda_{L''} + R_{L''}\Sigma_{\langle L''}$$

(3.46)
$$\begin{aligned}\Sigma_{L''\}} &= R_{L''}\Sigma_{\langle L''\}} \\ &= \Sigma_{L''\rangle}\Sigma_{L}^{-1}\begin{bmatrix}\Sigma_{\langle M} \\ \Sigma_{\{M}\end{bmatrix}\end{aligned}$$

where $\Sigma_{\{M} = \Sigma_{M\}}^{t}$, obtaining $\Sigma_{L'}$.

<u>Step 5:</u>

$$\Sigma_{M'\rangle} = R_{M'}\Sigma_{L\vee M}$$

$$\Sigma_{M'M'} = \Lambda_{M'} + R_{M'}\Sigma_{\langle M'}$$

(3.47)
$$\begin{aligned}\Sigma_{M'\}} &= R_{M'}\Sigma_{\langle M'\}} \\ &= \Sigma_{M'\rangle}\Sigma_{L\vee M}^{-1}\begin{bmatrix}\Sigma_{\langle L''} \\ \Sigma_{\{L''}\end{bmatrix}\end{aligned}$$

where $\Sigma_{\{L''} = \Sigma_{L''\}}^{t}$. Thus $\mathbf{P}_{\mathscr{K}}(I)$ consists of all $\Sigma \in \mathbf{P}(I)$ of the form

(3.48)
$$\Sigma = \left[\begin{array}{ccc:c:c} \Sigma_{L\wedge M} & \Sigma_{\langle L} & \Sigma_{\langle M} & \Sigma_{\langle L''} & \\ \Sigma_{L\rangle} & \Sigma_{LL} & \Sigma_{\{M} & & \Sigma_{\langle M'} \\ \cdashline{4-4} \Sigma_{M\rangle} & \Sigma_{M\}} & \Sigma_{MM} & \Sigma_{\{L''} & \\ \hdashline \multicolumn{2}{c:}{\Sigma_{L''\rangle}} & \Sigma_{L''\}} & \Sigma_{L''L''} & \Sigma_{\{M'} \\ \hdashline \multicolumn{3}{c:}{\Sigma_{M'\rangle}} & \Sigma_{M'\}} & \Sigma_{M'M'} \end{array}\right]$$

partitioned according to the decomposition

$$(3.49) \qquad I = D_{L\wedge M} \dot{\cup}\, D_{L} \dot{\cup}\, D_{M} \dot{\cup}\, D_{L''} \dot{\cup}\, D_{M'},$$

where $\Sigma_{M\}}$, $\Sigma_{L''\}}$, $\Sigma_{M'\}}$ satisfy (3.36), (3.46), (3.47), respectively. The precision matrix $\Lambda = \Sigma^{-1}$ satisfies the following three conditions: its $D_{M'} \times D_{L''}$ and $D_{L''} \times D_{M'}$ blocks are zero matrices, the $D_{L''} \times D_{M}$ and $D_{M} \times D_{L''}$ blocks of $\Sigma_{L'}^{-1}$ are zero matrices, and $\Sigma_{L\vee M}^{-1}$ has the form (3.39). Neither $\mathbf{P}_{\mathcal{K}}(I)$ nor $\mathbf{P}^{\mathcal{K}}(I)$ are linear. □

<u>Example 3.9.</u> Now consider the lattice $\mathcal{K}$ in Figure 3.9a:

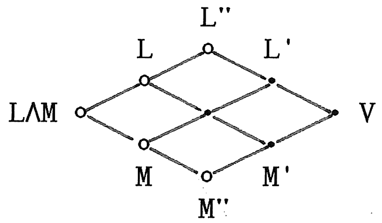

Figure 3.9a.

Although this lattice properly contains the lattices in Examples 3.7 and 3.8 as sublattices, the set $\mathbf{P}_{\mathcal{K}}(I)$ that it determines is much simpler than those in Examples 3.7 and 3.8. The reader may verify that $\mathbf{P}_{\mathcal{K}}(I)$ is identical to $\mathbf{P}_{\mathcal{M}}(I)$, where $\mathcal{M}$ is the sublattice in Figure 3.9b:

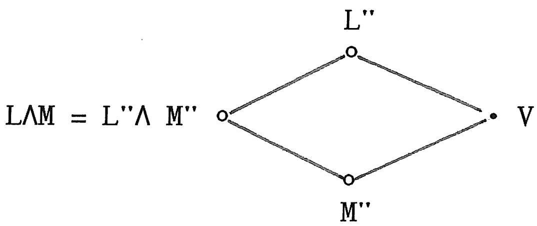

Figure 3.9b

The lattice $\mathcal{M}$

(compare to Example 3.5). Likewise, $P_{\mathcal{K}'}(I) = P_{\mathcal{M}'}(I)$ and $P_{\mathcal{K}''}(I) = P_{\mathcal{M}''}(I)$, where $\mathcal{K}'$ and $\mathcal{M}'$ are the sublattices in Figures 3.10a and 3.10b respectively:

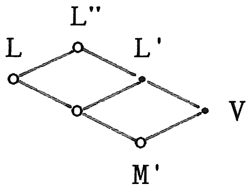

Figure 3.10a

The lattice $\mathcal{K}'$

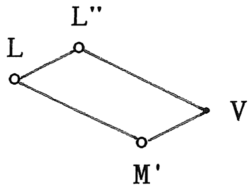

Figure 3.10b

The lattice $\mathcal{M}'$

and where $\mathcal{K}''$ and $\mathcal{M}''$ are the sublattices in Figures 3.11a and 3.11b respectively:

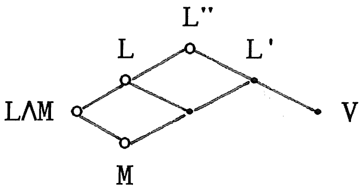

Figure 3.11a

The lattice $\mathcal{K}''$

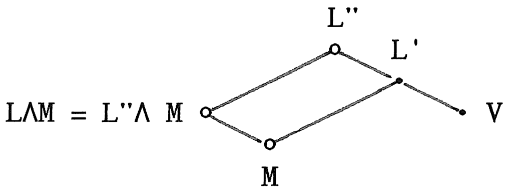

Figure 3.11b

The lattice $\mathcal{M}''$

□

<u>Remark 3.7.</u> It follows from the duality between the rings $\mathcal{R}$ and $\mathcal{R}'$ (cf. Remark 3.2) that for each finite distributive lattice $\mathcal{K} \subseteq \mathcal{K}(V)$ there exists an anti-isomorphic lattice $\mathcal{K}' \subseteq \mathcal{K}(V)$ such that $P_{\mathcal{K}'}(I) = P^{\mathcal{K}}(I)$ and $P^{\mathcal{K}'}(I) = P_{\mathcal{K}}(I)$. For example, if $\mathcal{K}$ is the lattice in Figure 3.4, then $\mathcal{K}'$ has the same form as the lattice in Figure 3.5, and conversely. □

§4. MAPPINGS THAT PRESERVE A DISTRIBUTIVE LATTICE $\mathcal{K}$ OF QUOTIENT SPACES.

4.1. The $\mathcal{K}$-preserving mappings: invariant formulation.

If as in Example 3.2, $\mathcal{K} = \{K_1 < \cdots < K_n = V\}$ is an ascending chain of quotients in V, then the corresponding kernels $p_1^{-1}(0) \supset \cdots \supset p_n^{-1}(0) = \{0\}$ form a descending chain of subspaces in V. It is well known that a linear mapping $f: V \to V$ preserves each $p_i^{-1}(0)$ if and only if the matrix of f wrt a basis adapted to $\mathcal{K}$ is lower block triangular. We call such a mapping $\mathcal{K}$-preserving. Furthermore, the group $\mathbf{GL}_{\mathcal{K}}(V)$ of all nonsingular $\mathcal{K}$-preserving mappings acts transitively on $\mathbf{P}^{\mathcal{K}}(V) = \mathbf{P}(V)$ under the usual action (cf. (4.1) and (4.11)). This group plays an important role in multivariate statistical analysis (cf. Eaton (1983), Andersson, Marden, and Perlman (1988)).

For a general distributive lattice $\mathcal{K} \subseteq \mathcal{K}(V)$, we define a linear mapping $f: V \to V$ to be $\mathcal{K}$-preserving if $f(p_K^{-1}(0)) \subseteq p_K^{-1}(0)$ for every $K \in \mathcal{K}$. The set of all matrices (wrt to a $\mathcal{K}$-adapted basis) for all $\mathcal{K}$-preserving mappings is a natural generalization of the class of (lower) block-triangular matrices (cf. (4.4) and Remark 4.3). It is shown in Proposition 4.1 that the group $\mathbf{GL}_{\mathcal{K}}(V)$ of nonsingular $\mathcal{K}$-preserving mappings acts transitively and properly on $\mathbf{P}^{\mathcal{K}}(V)$ under the action (4.1), an important fact for the analysis of the testing problem in Section 6.

The class $\mathbf{End}_{\mathcal{K}}(V)$ of all $\mathcal{K}$-preserving mappings of V to V is a subalgebra of the algebra $\mathbf{End}(V)$ of all linear mappings of V to V and $\mathbf{GL}_{\mathcal{K}}(V)$ is a subgroup of the general linear group $\mathbf{GL}(V)$ of V. Trivially $\mathbf{End}_{\{V\}}(V) = \mathbf{End}(V)$ and $\mathbf{GL}_{\{V\}}(V) = \mathbf{GL}(V)$.

Remark 4.1. If $V = \mathbb{R}^I$ where I is a finite set, then we identify the algebra $\mathbf{End}(\mathbb{R}^I)$ with the matrix algebra $\mathbf{M}(I) := \mathbf{M}(I\times I)$ and the group $\mathbf{GL}(V)$ with $\mathbf{M}^*(I)$, the group of all nonsingular $I\times I$ matrices in the usual way. If $\mathcal{R}$ is a ring of subsets of I such that $I \in \mathcal{R}$ (cf. Example 1.1) then under this identification, $\mathbf{End}_{\mathcal{K}(\mathcal{R})}(\mathbb{R}^I)$ becomes a subalgebra $\mathbf{M}_{\mathcal{R}}(I)$ of $\mathbf{M}(I)$ and $\mathbf{GL}_{\mathcal{K}(\mathcal{R})}(\mathbb{R}^I)$ becomes a subgroup $\mathbf{M}^*_{\mathcal{R}}(I)$ of $\mathbf{M}^*(I)$. □

4.2. Transitive action: invariant formulation.

Proposition 4.1. The action

$$\text{(4.1)} \qquad \begin{aligned} \mathbf{GL}_{\mathcal{K}}(V)\times\mathbf{P}^{\mathcal{K}}(V) &\to \mathbf{P}^{\mathcal{K}}(V) \\ (f,\delta) &\to \delta\circ(f^{-1}\times f^{-1}) \end{aligned}$$

is well-defined, transitive, continuous, and proper.

Proof: If $f \in \mathbf{GL}_{\mathcal{K}}(V)$ and $\delta \in \mathbf{P}^{\mathcal{K}}(V)$, then the subspaces $\{p_K^{-1}(0)\,|K\in\mathcal{K}\}$ are mutually geometrically orthogonal (GO) wrt $\delta\circ(f^{-1}\times f^{-1})$, hence $\delta\circ(f^{-1}\times f^{-1}) \in \mathbf{P}^{\mathcal{K}}(V)$ by Remark 2.6. Thus (4.1) is well-defined. Next let $(e_i'\,|i\in I)$ and $(e_i''\,|i\in I)$ be two bases for V adapted to $\mathcal{K}$. If $f \in \mathbf{GL}(V)$ is such that $f(e_i') = e_i''$, $i \in I$, then $f \in \mathbf{GL}_{\mathcal{K}}(V)$. Furthermore, it follows from Theorem 3.2 that for any $\delta \in \mathbf{P}^{\mathcal{K}}(V)$ there exists a basis $(e_i\,|i\in I)$ adapted to $\mathcal{K}$ such that the matrix for δ wrt $(e_i\,|i\in I)$ is the $I\times I$ identity matrix. It follows readily from these two facts that the action (4.1) is transitive.

Since $\mathbf{GL}_{\mathcal{K}}(V)$ is a closed subgroup of the locally compact group $\mathbf{GL}(V)$ it is locally compact in the relative topology. Furthermore, $\mathbf{P}^{\mathcal{K}}(V)$ is a closed subset of the locally compact space $P(V)$, hence is also locally compact. Thus the action (4.1) is the restriction of the classical action

$$\begin{aligned} GL(V)\times P(V) &\to P(V) \\ (f,\delta) &\to \delta\circ(f^{-1}\times f^{-1}) \end{aligned} \tag{4.2}$$

to the closed subset $GL_{\mathcal{H}}(V)\times P^{\mathcal{H}}(V)$. Since it is well known that the action (4.2) is continuous and proper (cf. [A] (1982)), the action (4.1) inherits these properties. □

Remark 4.2. By Proposition 4.1, the action

$$\begin{aligned} GL_{\mathcal{H}}(V)\times P_{\mathcal{H}}(V) &\to P_{\mathcal{H}}(V) \\ (f,\sigma) &\to \sigma\circ(f^{t}\times f^{t}). \end{aligned} \tag{4.3}$$

induced on $P_{\mathcal{H}}(V)$ by (4.1) is also transitive, continuous, and proper. □

4.3. The $\mathcal{H}$-preserving mappings: matrix formulation.

If a basis $(e_i \mid i\in I)$ is choosen for V, then End(V) and GL(V) are identified with $\mathbf{M}(I)$ and $\mathbf{M}^*(I)$, respectively, in the usual way. If this basis is adapted $\mathcal{H}$, then under this identification $\mathrm{End}_{\mathcal{H}}(V)$ and $GL_{\mathcal{H}}(V)$ are identified with $\mathbf{M}_{\mathfrak{R}(\mathcal{H})}(I) =: \mathbf{M}_{\mathcal{H}}(I)$ and $\mathbf{M}^*_{\mathfrak{R}(\mathcal{H})}(I) =: \mathbf{M}^*_{\mathcal{H}}(I)$, resp. (cf. Remarks 2.2 and 4.1).

For any $A \in \mathbf{M}(I)$ and $K \in \mathcal{H}$ let A_K denote the $I_K\times I_K$ submatrix of A. For $L,M \in J(\mathcal{H})$ let A_{LM} denote the $D_L\times D_M$ submatrix of A, let $A_{L\rangle}$ denote the $D_L\times I_{J(L)}$ submatrix of A, and let $A_{\langle L\rangle} = A_{J(L)}$.

Proposition 4.2. The matrix algebra $\mathbf{M}_{\mathcal{K}}(I)$ is characterized as follows:

(4.4) $$\mathbf{M}_{\mathcal{K}}(I) = \{A \in \mathbf{M}(I) \mid \forall L, M \in J(\mathcal{K}) : A_{LM} \neq 0 \Rightarrow M \leq L\}.$$

Proof: First note that for every $L \in \mathcal{K}$,

(4.5) $$x \in p_L^{-1}(0) \Leftrightarrow (\forall K \in J(\mathcal{K}) : K \leq L \Rightarrow x_{[K]} = 0)$$

(cf. (3.15)). For fixed $M \in J(\mathcal{K})$ let ϵ denote an arbitrary column vector in $\mathbb{R}^I$ satisfying $\epsilon_{[K]} = 0$ for $K \in J(\mathcal{K})$, $K \neq M$. Then for every $L \in J(\mathcal{K})$ with $M \nleq L$, one has that $\epsilon \in p_L^{-1}(0)$, hence $A\epsilon \in p_L^{-1}(0)$ for every $A \in \mathbf{M}_{\mathcal{K}}(I)$. In particular, by (4.5),

$$0 = (A\epsilon)_{[L]} = A_{LM}\epsilon_{[M]},$$

hence $A_{LM} = 0$ since $\epsilon_{[M]}$ is arbitrary.

Conversely, if $A \in \mathbf{M}(I)$ satisfies condition (4.4), then for every $L \in \mathcal{K}$, every $x \in p_L^{-1}(0)$, and every $K \in J(\mathcal{K})$ such that $K \leq L$,

$$(Ax)_{[K]} = \Sigma(A_{KM}x_{[M]} \mid M \in J(\mathcal{K})) = \Sigma(A_{KM}x_{[M]} \mid M \in J(\mathcal{K}),\ M \leq K) = 0$$

The second equality follows from (4.4), while the third equality holds since $M \leq K \leq L$ and $M \in J(\mathcal{K})$ imply that $x_{[M]} = 0$, by (4.5). Thus again by (4.5), $Ax \in p_L^{-1}(0)$, hence $A \in \mathbf{M}_{\mathcal{K}}(I)$. □

Remark 4.3. Let $O_{\mathcal{K}} = K_1, K_2, \cdots, K_{|J(\mathcal{K})|}$ be a never-decreasing listing of the members of the poset $J(\mathcal{K})$ as in Remark 3.6. If every $A \in \mathbf{M}(I)$ is partitioned according to the ordered decomposition

(4.6) $$I = D_1 \dot{\cup}\, D_2 \dot{\cup} \cdots \dot{\cup}\, D_{|J(\mathcal{K})|}, \qquad (D_k \equiv D_{K_k})$$

then it is seen from (4.4) that $\mathbf{M}_{\mathcal{K}}(I)$ becomes a subalgebra of the algebra of lower block triangular matrices in the usual sense. □

Remark 4.4. For $A \in \mathbf{M}_{\mathcal{K}}(I)$, $x \in \mathbb{R}^I$, $L \in \mathcal{K}$, and $K \in J(\mathcal{K})$,

(4.7) $$(Ax)_L = A_L x_L$$

(4.8) $$(Ax)_{[K]} = A_{KK} x_{[K]} + A_{K\rangle} x_{\langle K\rangle}.$$ □

Remark 4.5. The linear mapping

(4.9) $$\begin{aligned} \mathbf{M}_{\mathcal{K}}(I) &\to \times(\mathbf{M}(D_K \times I_{J(K)}) \times \mathbf{M}(D_K) \mid K \in J(\mathcal{K})) \\ A &\to ((A_{K\rangle}, A_{KK}) \mid K \in J(\mathcal{K})) \end{aligned}$$

is bijective, since for every $K \in J(\mathcal{K})$, those entries in the $D_K \times (I \setminus D_K)$ submatrix of A that do not lie in $A_{K\rangle}$ must be zero. Under the correspondence (4.9) the subset $\mathbf{M}^*_{\mathcal{K}}(I)$ corresponds to the subset

(4.10) $$\times(\mathbf{M}(D_K \times I_{J(K)}) \times \mathbf{M}^*(D_K) \mid K \in J(\mathcal{K})).$$ □

4.4. Transitive action: matrix formulation.

In matrix formulation the action (4.1) assumes the form

$$(4.11)\qquad \begin{aligned} \mathbf{M}^*_{\mathcal{K}}(I)\times\mathbf{P}^{\mathcal{K}}(I) &\to \mathbf{P}^{\mathcal{K}}(I) \\ (A,\Lambda) &\to (A^{-1})^t\Lambda A^{-1}, \end{aligned}$$

while (4.3) becomes

$$(4.12)\qquad \begin{aligned} \mathbf{M}^*_{\mathcal{K}}(I)\times\mathbf{P}_{\mathcal{K}}(I) &\to \mathbf{P}_{\mathcal{K}}(I) \\ (A,\Sigma) &\to A\Sigma A^t. \end{aligned}$$

Remark 4.6. Since both $\mathbf{P}^{\mathcal{K}}(I)$ and $\mathbf{P}_{\mathcal{K}}(I)$ contains the I×I identity matrix, it follows from the transitivity of the actions (4.11) and (4.12) that

$$(4.13)\qquad \mathbf{P}^{\mathcal{K}}(I) = \{A^tA\in\mathbf{P}(I)\,|\,A\in\mathbf{M}^*_{\mathcal{K}}(I)\}$$

$$(4.14)\qquad \mathbf{P}_{\mathcal{K}}(I) = \{AA^t\in\mathbf{P}(I)\,|\,A\in\mathbf{M}^*_{\mathcal{K}}(I)\}. \qquad \square$$

If $\mathcal{K} = \{V\}$ then $\mathbf{M}^*_{\mathcal{K}}(I) = \mathbf{M}^*(I)$ and $\mathbf{P}^{\mathcal{K}}(I) = \mathbf{P}_{\mathcal{K}}(I) = \mathbf{P}(I)$, so both actions (4.11) and (4.12) reduce to the well-known transitive action of $\mathbf{M}^*(I)$ on $\mathbf{P}(I)$. If $\mathcal{K}$ is a chain as in Examples 3.1 or 3.2, then again $\mathbf{P}^{\mathcal{K}}(I) = \mathbf{P}_{\mathcal{K}}(I) = \mathbf{P}(I)$, but now $\mathbf{M}^*_{\mathcal{K}}(I)$ is a group of nonsingular lower block-triangular matrices in the usual sense, and the actions (4.11) and (4.12) are the well-known transitive actions of $\mathbf{M}^*_{\mathcal{K}}(I)$ on $\mathbf{P}(I)$.

Now consider the lattices $\mathcal{K}$ in Examples 3.3–3.9, respectively. By (4.4), in these seven examples the corresponding matrix algebras $\mathbf{M}_{\mathcal{K}}(I)$ consist of all I×I matrices of the following forms:

(4.15a)
$$A = \begin{bmatrix} A_L & 0 \\ 0 & A_M \end{bmatrix}$$

(4.15b)
$$A = \left[\begin{array}{cc|c} A_L & 0 & 0 \\ 0 & A_M & 0 \\ \hline \multicolumn{2}{c|}{A_{V>}} & A_{VV} \end{array}\right]$$

(4.15c)
$$A = \begin{bmatrix} A_{L\wedge M} & 0 & 0 \\ A_{L>} & A_{LL} & 0 \\ A_{M>} & 0 & A_{MM} \end{bmatrix}$$

(4.15d)
$$A = \left[\begin{array}{ccc|c} A_{L\wedge M} & 0 & 0 & 0 \\ A_{L>} & A_{LL} & 0 & 0 \\ A_{M>} & 0 & A_{MM} & 0 \\ \hline \multicolumn{3}{c|}{A_{V>}} & A_{VV} \end{array}\right]$$

(4.15e)
$$A = \left[\begin{array}{ccc|cc} A_{L\wedge M} & 0 & 0 & 0 & 0 \\ A_{L>} & A_{LL} & 0 & 0 & 0 \\ A_{M>} & 0 & A_{MM} & 0 & 0 \\ \hline \multicolumn{3}{c|}{A_{L'>}} & A_{L'L'} & 0 \\ \multicolumn{3}{c|}{A_{M'>}} & 0 & A_{M'M'} \end{array}\right]$$

(4.15f)
$$A = \left[\begin{array}{ccc|cc} A_{L\wedge M} & 0 & 0 & 0 & 0 \\ A_{L>} & A_{LL} & 0 & 0 & 0 \\ A_{M>} & 0 & A_{MM} & 0 & 0 \\ \hline \multicolumn{2}{c|}{A_{L''>}} & 0 & A_{L''L''} & 0 \\ \cline{1-3} \multicolumn{3}{c|}{A_{M'>}} & 0 & A_{M'M'} \end{array}\right]$$

(4.15g)
$$A = \left[\begin{array}{ccc:cc} A_{L\wedge M} & 0 & 0 & 0 & 0 \\ A_{L>} & A_{LL} & 0 & 0 & 0 \\ A_{M>} & 0 & A_{MM} & 0 & 0 \\ \hdashline \multicolumn{2}{c:}{A_{L''>}} & 0 & A_{L''L''} & 0 \\ \hdashline A_{M'',L\wedge M} & 0 & A_{M'',M} & 0 & A_{M''M''} \end{array}\right].$$

(Note that $(A_{M'',L\wedge M} \vdots A_{M'',M}) = A_{M''>}$ in (4.15g).) The reader may verify directly from (4.15a)-(4.15g) that (4.13) and (4.14) hold in these seven examples.

The action induced by (4.12) on the $\mathcal{K}$-parametrization of $\mathbf{P}_{\mathcal{K}}(I)$ in (3.13) is the following:

(4.16)
$$\begin{gathered} \mathbf{M}^*_{\mathcal{K}}(I)\times(X(\mathbf{M}(D_K\times I_{J(K)})\times\mathbf{P}(D_K)\,|K\in J(\mathcal{K}))) \\ \to X(\mathbf{M}(D_K\times I_{J(K)})\times\mathbf{P}(D_K)\,|K\in J(\mathcal{K})) \\ (A\ ,\ ((\Sigma_{K>}\Sigma^{-1}_{<K>},\ \Sigma_{K\cdot})\,|K\in J(\mathcal{K}))) \\ \to((A_{KK}\Sigma_{K>}\Sigma^{-1}_{<K>}A^{-1}_{<K>} + A_{K>}A^{-1}_{<K>},\ A_{KK}\Sigma_{K\cdot}A^t_{KK})\,|K\in J(\mathcal{K}))) \end{gathered}$$

This fact is needed in Section 6, and also for the proof of the following lemma.

<u>Lemma 4.1.</u> For $\Sigma \in \mathbf{P}_{\mathcal{K}}(I)$,

(4.17)
$$\det(\Sigma) = \Pi(\det(\Sigma_{K\cdot})\,|K\in J(\mathcal{K})).$$

<u>Proof:</u> By (4.14), there exists $A \in \mathbf{M}^*_{\mathcal{K}}(I)$ such that $AA^t = \Sigma$. It follows from (4.16) that $A_{KK}A^t_{KK} = \Sigma_{K\cdot}$, $K \in J(\mathcal{K})$. Thus by Remark 4.3,

$$\begin{aligned}\det(\Sigma) &= \det(AA^t)\\ &= \Pi(\det(A_{KK}A_{KK}^t)\,|K\in J(\mathcal{K}))\\ &= \Pi(\det(\Sigma_{K\bullet})\,|K\in J(\mathcal{K})).\end{aligned}$$ □

§5. LIKELIHOOD INFERENCE FOR A NORMAL MODEL DETERMINED BY PAIRWISE CONDITIONAL INDEPENDENCE.

Although it is always desirable to describe and analyze a statistical model in an invariant (coordinate-free) way, the analysis in this section is presented in matrix formulation so that the reader may see its relation to classical multivariate analysis more easily. Nevertheless, it is important to note that the models and their analysis may be described in an invariant manner.

The multivariate normal distribution on V with mean 0 and covariance $\sigma \in \mathbf{P}(V^*)$ is denoted by $N(\sigma)$. For a distributive lattice $\mathcal{K} \subseteq \mathcal{K}(V)$, the normal statistical model $\mathcal{N}_V(\mathcal{K})$ determined by pairwise conditional independence (CI) with respect to $\mathcal{K}$ is defined to be

(5.1) $$\mathcal{N}_V(\mathcal{K}) = (N(\sigma)\,|\sigma\in\mathbf{P}_{\mathcal{K}}(V)).$$

If $x \in V$ represents an observation from the model (5.1), thus the model states that for every $L,M \in \mathcal{K}$, x_L is CI of x_M given $x_{L\wedge M}$, which is written as follows:

(5.2) $$x_L \perp\!\!\!\perp x_M\,|x_{L\wedge M}.$$

The model $\mathcal{N}_V(\mathcal{K})$ is a curved exponential family; it is linear if and only if $P^{\mathcal{K}}(V)$ is a linear set, i.e., closed under positive linear combinations. In the linear case the ML estimator $\hat{\sigma}$ based on N i.i.d. observations from $\mathcal{N}_V(\mathcal{K})$ is a minimal sufficient statistic, but is not necessarily sufficient in the general case.

If $(e_i|i\in I)$ is a basis for V adapted to $\mathcal{K}$, then V may be identified with $\mathbb{R}^I$ and the model (5.1) may be expressed in matrix formulation as

$$(5.3) \qquad \mathcal{N}_I(\mathcal{K}) := (N(\Sigma)|\Sigma\in \mathbf{P}_{\mathcal{K}}(I)),$$

where $N(\Sigma)$ denotes the normal distribution on $\mathbb{R}^I$ with covariance matrix Σ.

5.1. Maximum likelihood estimation.

Consider N independent and identically distributed (i.i.d.) observations $x_1,\dots,x_N$ from the model (5.3) and set

$$(5.4) \qquad y := (x_1,\dots,x_N) \in \mathbf{M}(I\times\{1,\dots,N\}) =: \mathbf{M}(I\times N).$$

For $L \in \mathcal{K}$ let y_L denote the $I_L\times N$ submatrix of y, while for $K \in J(\mathcal{K})$ partition y_K according to (2.4) as follows:

$$y_K = \begin{pmatrix} y_{\langle K\rangle} \\ y_{[K]} \end{pmatrix}.$$

By (3.14) and (4.17), the likelihood function for this statistical model is given by the mapping

$$(5.5)\qquad \mathbf{P}_{\mathcal{K}}(I)\times\mathbf{M}(I\times N) \to \mathbb{R}$$

$$(\Sigma,y) \to (\det(\Sigma))^{-N/2}\exp(-\mathrm{tr}(\Sigma^{-1}yy^t)/2)$$

$$= \Pi((\det(\Sigma_{K\cdot}))^{-N/2}\times\exp(\mathrm{tr}(\Sigma_{K\cdot}^{-1}(y_{[K]}-\Sigma_{K\rangle}\Sigma_{\langle K\rangle}^{-1}y_{\langle K\rangle})(\cdots)^t)/2)\,|K\in J(\mathcal{K})).$$

Note that the factor corresponding to $K \in J(\mathcal{K})$ is the density for the conditional distribution of $y_{[K]}$ given $y_{\langle K\rangle}$.

It now follows readily from (3.13) and well-known results for the multivariate normal linear regression model that the maximum likelihood (ML) estimator $\hat{\Sigma}(y) \in \mathbf{P}_{\mathcal{K}}(I)$ for $\Sigma \in \mathbf{P}_{\mathcal{K}}(I)$ is unique if it exists, and it exists for a.e. $y \in \mathbf{M}(I\times N)$ if and only if

$$(5.6)\qquad N \geq \max\{|I_{J(K)}|+|D_K|\ \ |K\in J(\mathcal{K})\} = \max\{|I_K|\ \ |K\in J(\mathcal{K})\}.$$

(Note that $\dim(K) = |I_K|$, $K \in \mathcal{K}$.) In this case the $\mathcal{K}$-parameters of $\hat{\Sigma}$ are determined from the usual formulas for regression estimators:

$$(5.7)\qquad \widehat{\Sigma_{K\rangle}\Sigma_{\langle K\rangle}^{-1}} = S_{K\rangle}S_{\langle K\rangle}^{-1},\quad N\hat{\Sigma}_{K\cdot} = S_{K\cdot},\qquad K \in J(\mathcal{K}),$$

where $S(y) = yy^t$ is the empirical covariance matrix. The explicit expression for $\hat{\Sigma}$ itself may be obtained from its $\mathcal{K}$-parameters in (5.7) by means of the reconstruction algorithm given in Remark 3.6.

If $V \in J(\mathcal{K})$ then the condition (5.6) reduces to $N \geq |I|$, so in this case S is positive definite a.e., hence a fortiori $S_{K\cdot}$ exists and is positive definite for every $K \in J(\mathcal{K})$. If, on the other hand, $V \notin J(\mathcal{K})$, then condition (5.6) does not guarantee that S is positive definite, but it still guarantees that $S_{K\cdot}$ exists and is positive definite a.e.

By Lemma 4.1, when (5.6) is satisfied the maximum value of the likelihood function (5.5) is given by

$$(5.8) \qquad c\cdot\Pi((\det(\hat{\Sigma}_{K\cdot}))^{-N/2}|K\in J(\mathcal{K})) = c\cdot(\det(\hat{\Sigma}))^{-N/2}$$

where $c = N^{N/2}\times\exp(-N|I|/2)$. This fact is used in Section 6 to express the likelihood ratio statistic for testing one model against another.

5.2. Examples of pairwise conditional independence models.

For each lattice $\mathcal{K}$ in Examples 3.1–3.9, consider the normal statistical model $\mathcal{N}_I(\mathcal{K})$. When $\mathcal{K}$ is a chain as in Examples 3.1 and 3.2, $P_{\mathcal{K}}(I) = P(I)$ and $\mathcal{N}_I(\mathcal{K})$ is the unrestricted covariance model irrregardless of the length of the chain. (The $\mathcal{K}$-parametrization of $\mathbf{P}_{\mathcal{K}}(I)$ does depend on this length, however.) Condition (5.6) for existence of the ML estimator $\hat{\Sigma}$ reduces to the familar condition $N \geq |I|$, while (5.7) reduces to $N\hat{\Sigma} = S$.

For the lattice $\mathcal{K}$ in Example 3.3, partition the observation $x \in \mathbb{R}^I$ according to (3.30) as $x = (x_L^t, x_M^t)^t$. The model $\mathcal{N}_I(\mathcal{K})$ states simply that $x_L \perp\!\!\!\perp x_M$. According to (5.6), the ML estimator $\hat{\Sigma}$ exists if and only if $N \geq \max\{|I_L|, |I_M|\}$ (whereas S is positive definite if and only if $N \geq |I|$) and is given by $N\hat{\Sigma} = \mathrm{Diag}(S_L, S_M)$.

For the lattice $\mathcal{K}$ in Example 3.4, partition $x \in \mathbb{R}^I$ according to (3.33) as $x = (x_L^t, x_M^t, x_{[V]}^t)^t$. Then the model $\mathcal{N}_I(\mathcal{K})$ again states that $x_L \perp\!\!\!\perp x_M$. Condition (5.6) for the existence of the ML estimator takes the form $N \geq |I|$, while from (5.7),

$$N\hat{\Sigma}_L = S_L, \quad N\hat{\Sigma}_M = S_M, \quad \widehat{\Sigma_{V>}\Sigma_{LVM}^{-1}} = S_{V>}S_{LVM}^{-1}, \quad N\hat{\Sigma}_{V\cdot} = S_{V\cdot}.$$

Finally, we may reconstruct $\hat{\Sigma}$ from its $\mathcal{H}$-parameters by following Steps 1-4 in Example 3.4 to obtain

$$N\hat{\Sigma}_{LVM} = \mathrm{Diag}(S_L, S_M)$$
$$N\hat{\Sigma}_{V>} = S_{V>}S_{LVM}^{-1}\mathrm{Diag}(S_L, S_M)$$
$$N\hat{\Sigma}_{VV} = S_{V\cdot} + S_{V>}(\mathrm{Diag}(S_L, S_M))^{-1}S_{<V} \ (\neq S_{VV}).$$

In Example 3.5, $x \in \mathbb{R}^I$ is partitioned according to (3.38) as $(x_{L\wedge M}^t, x_{[L]}^t, x_{[M]}^t)^t$. The model $\mathcal{N}_I(\mathcal{H})$ states that $x_{[L]} \perp\!\!\!\perp x_{[M]} \mid x_{L\wedge M}$. Condition (5.6) becomes $N \geq \max\{|I_L|, |I_M|\}$, while (5.7) becomes

(5.9a) $$N\hat{\Sigma}_{L\wedge M} = S_{L\wedge M}$$
(5.9b) $$\widehat{\Sigma_{L>}\Sigma_{L\wedge M}^{-1}} = S_{L>}S_{L\wedge M}, \quad N\hat{\Sigma}_{L\cdot} = S_{L\cdot}$$
(5.9c) $$\widehat{\Sigma_{M>}\Sigma_{L\wedge M}^{-1}} = S_{M>}S_{L\wedge M}, \quad N\hat{\Sigma}_{M\cdot} = S_{M\cdot}.$$

By Steps 1-4 in Example 3.5, $\hat{\Sigma}$ is given by (5.9a) and

(5.10a) $$N\hat{\Sigma}_{L>} = S_{L>}, \quad N\hat{\Sigma}_{LL} = S_{LL}$$
(5.10b) $$N\hat{\Sigma}_{M>} = S_{M>}, \quad N\hat{\Sigma}_{MM} = S_{MM}$$
(5.10c) $$N\hat{\Sigma}_{M\}} = S_{M>}S_{L\wedge M}^{-1}S_{<L} \ (\neq S_{M\}}).$$

In Example 3.6, $x \in \mathbb{R}^I$ is partitioned as $(x_{L\wedge M}^t, x_{[L]}^t, x_{[M]}^t, x_{[V]}^t)^t$ and the model $\mathcal{N}_I(\mathcal{H})$ states that $x_{[L]} \perp\!\!\!\perp x_{[M]} \mid x_{L\wedge M}$. Condition (5.6) reduces to $N \geq |I|$, while (5.7) is given by (5.9a,b,c) and

$$\widehat{\Sigma_{V>}\Sigma_{LVM}^{-1}} = S_{V>}S_{LVM}^{-1}, \quad N\hat{\Sigma}_{V\cdot} = S_{V\cdot}.$$

From Steps 1-4 in Example 3.6, $\hat{\Sigma}$ is given by (5.9a), (5.10a,b,c), and

$$N\hat{\Sigma}_{V>} = S_{V>}S_{LVM}^{-1}(N\hat{\Sigma}_{LVM}) \quad (\neq S_{V>})$$
$$N\hat{\Sigma}_{VV} = S_{V\cdot} + S_{V>}S_{LVM}^{-1}(N\hat{\Sigma}_{LVM})S_{LVM}^{-1}S_{<V} \quad (\neq S_{VV})$$

where, from (5.9a) and (5.10a,b,c),

(5.11)
$$N\hat{\Sigma}_{LVM} = \begin{bmatrix} S_{L\wedge M} & S_{<L} & S_{<M} \\ S_{L>} & S_{LL} & S_{L>}S_{L\wedge M}^{-1}S_{<M} \\ S_{M>} & S_{M>}S_{L\wedge M}^{-1}S_{<L} & S_{MM} \end{bmatrix}.$$

In Example 3.7, $x \in \mathbb{R}^I$ is partitioned as $(x_{L\wedge M}^t, x_{[L]}^t, x_{[M]}^t, x_{[L']}^t, x_{[M']}^t)^t$ and the model $\mathcal{N}_I(\mathcal{H})$ states that

$$x_{[L]} \perp\!\!\!\perp x_{[M]} \mid x_{L\wedge M} \text{ and that } x_{[L']} \perp\!\!\!\perp x_{[M']} \mid (x_{L\wedge M}, x_{[L]}, x_{[M]}).$$

(Note that $x_{L'\wedge M'} = x_{LVM} = (x_{L\wedge M}, x_{[L]}, x_{[M]})$.) Condition (5.6) becomes $N \geq \max\{|I_{L'}|, |I_{M'}|\}$, while (5.7) is given by (5.9a,b,c) and (5.9b,c) with L,M replaced by L',M' (note that $S_{L'\wedge M'} = S_{LVM}$). From Steps 1-5 in Example 3.7, $\hat{\Sigma}$ is given by (5.9a), (5.10a,b,c), and

$$N\hat{\Sigma}_{L'>} = S_{L'>}S_{LVM}^{-1}(N\hat{\Sigma}_{LVM})$$
$$N\hat{\Sigma}_{L'L'} = S_{L'\cdot} + S_{L'>}S_{LVM}^{-1}(N\hat{\Sigma}_{LVM})S_{LVM}^{-1}S_{<L'}$$

(5.12a)
$$N\hat{\Sigma}_{M'>} = S_{M'>}S_{LVM}^{-1}(N\hat{\Sigma}_{LVM})$$

(5.12b)
$$N\hat{\Sigma}_{M'M'} = S_{M'\cdot} + S_{M'>}S_{LVM}^{-1}(N\hat{\Sigma}_{LVM})S_{LVM}^{-1}S_{<M'}$$

(5.12c)
$$N\hat{\Sigma}_{M'\}} = S_{M'>}S_{LVM}^{-1}(N\hat{\Sigma}_{LVM})S_{LVM}^{-1}S_{<L'} \quad (\neq S_{M'\}})$$

where $N\hat{\Sigma}_{L\vee M}$ is given by (5.11).

In Example 3.8, $x \in \mathbb{R}^I$ is partitioned as $(x^t_{L\wedge M}, x^t_{[L]}, x^t_{[M]}, x^t_{[L'']}, x^t_{[M']})^t$. It may be seen from the form (3.48) of $\Sigma \in P_{\mathcal{H}}(I)$ that the model $\mathcal{N}_I(\mathcal{H})$ is determined by the following three conditions:

$$\begin{aligned}
&\text{(i)} \quad x_{[L]} \perp\!\!\!\perp x_{[M]} \mid x_{L\wedge M} \\
&\text{(ii)} \quad x_{[M]} \perp\!\!\!\perp x_{[L'']} \mid (x_{L\wedge M}, x_{[L]}) \\
&\text{(iii)} \quad x_{[L'']} \perp\!\!\!\perp x_{[M']} \mid (x_{L\wedge M}, x_{[L]}, x_{[M]}).
\end{aligned}$$

Condition (5.6) becomes $N \geq \max\{|I_{L''}|, |I_{M'}|\}$, while (5.7) is given by (5.9a,b,c),

$$\widehat{\Sigma_{L''>}\Sigma_L^{-1}} = S_{L''>}S_L^{-1}, \qquad N\hat{\Sigma}_{L''\cdot} = S_{L''\cdot}$$
$$\widehat{\Sigma_{M'>}\Sigma_{L\vee M}^{-1}} = S_{M'>}S_{L\vee M}^{-1}, \qquad N\hat{\Sigma}_{M'\cdot} = S_{M'\cdot}.$$

From Steps 1-5 in Example 3.8, $\hat{\Sigma}$ is given by (5.9a) and (5.10a,b,c), by

$$N\hat{\Sigma}_{L''>} = S_{L''>}, \qquad N\hat{\Sigma}_{L''L''} = S_{L''L''}$$

$$N\hat{\Sigma}_{L''\not>} = S_{L''>}S_L^{-1}\begin{bmatrix} S_{<M} \\ S_{L>}S_{L\wedge M}^{-1}S_{<M} \end{bmatrix},$$

by (5.12a,b), and by

$$N\hat{\Sigma}_{M'\not>} = S_{M'>}S_{L\vee M}^{-1}\begin{bmatrix} S_{<L''} \\ N\hat{\Sigma}_{\not<L''} \end{bmatrix}.$$

Finally, for the lattice $\mathcal{K}$ in Example 3.9, x is partitioned as $(x^t_{L\wedge M}, x^t_{[L]}, x^t_{[M]}, x^t_{[L'']}, x^t_{[M'']})^t$. If one proceeds as above, one finds that the model $\mathcal{N}_I(\mathcal{K})$ is determined by the single condition that

$$(x_{[L]}, x_{[L'']}) \perp\!\!\!\perp (x_{[M]}, x_{[M'']}) \mid x_{L\wedge M}.$$

This reflects the fact that this model is of the same form as that in Example 3.5 (see the discussion in Example 3.9).

Remark 5.1. Recall the definition of the normal model $\mathcal{N}_I(\mathcal{K})$ for a distributive lattice $\mathcal{K}$: for every pair $L, M \in \mathcal{K}$, $x_L \perp\!\!\!\perp x_M \mid x_{L\wedge M}$. It may be seen from the above examples that many of these conditions are redundant and may be omitted, for example whenever $L \leq M$. More generally, if $L \leq L'$, $M \leq M'$, and $L \wedge M = L' \wedge M'$, the CI of $x_{L'}$ and $x_{M'}$ implies the CI of x_L and x_M, hence the latter condition may be omitted. The important question of characterizing a minimal set of CI conditions that determines $\mathcal{N}_I(\mathcal{K})$ is currently under investigation. For a given lattice $\mathcal{K}$, however, such minimal determining sets are not unique. In Example 3.8, for example, the following four sets of CI conditions are (equivalent) minimal determining sets for $\mathcal{N}_I(\mathcal{K})$:

(i) $x_L \perp\!\!\!\perp x_M \mid x_{L\wedge M}$; (ii) $x_{L\vee M} \perp\!\!\!\perp x_{L''} \mid x_L$; (iii) $x_{L'} \perp\!\!\!\perp x_{M'} \mid x_{L\vee M}$;

(i) $x_L \perp\!\!\!\perp x_M \mid x_{L\wedge M}$; (ii) $x_{L''} \perp\!\!\!\perp x_{M'} \mid x_L$;

(i) $x_M \perp\!\!\!\perp x_{L''} \mid x_{L\wedge M}$; (ii) $x_{L'} \perp\!\!\!\perp x_{M'} \mid x_{L\vee M}$;

(i) $x_M \perp\!\!\!\perp x_{L''} \mid x_{L\wedge M}$; (ii) $x_{L''} \perp\!\!\!\perp x_{M'} \mid x_L$. □

Remark 5.2. For $I = \{1,2,3,4\}$, consider the statistical model consisting of all normal distributions on $\mathbb{R}^I$ such that x_1 is independent of x_2 and x_3 is independent of x_4. It is readily seen that this model is not of the form $\mathcal{N}_I(\mathcal{K})$ for any $\mathcal{K}$. The same is true for the normal model determined by the two conditions that x_1 and x_2 are CI given (x_3,x_4) and x_3 and x_4 are CI given (x_1,x_2). □

Remark 5.3. The general model $\mathcal{N}_I(\mathcal{K})$ is defined by the pairwise CI requirement (5.2) for every pair $L,M \in \mathcal{K}$. This requirement does not necessary imply, however, that for every subset $\mathcal{S} \subseteq \mathcal{K}$, $(x_K|K\in\mathcal{S})$ are mutually CI given $x_{\wedge(K|K\in\mathcal{S})}$. For the lattice $\mathcal{K}$ in Example 3.9, this may be seen by considering the subset $\mathcal{S} = \{L'', L\vee M, M''\}$. □

Remark 5.4. An alternative statistical interpretation of the CI model $\mathcal{N}_I(\mathcal{K})$ may be obtained from (4.14): $x \equiv (x_{[K]}|K\in J(\mathcal{K})) \in \mathbb{R}^I$ is an observation from the normal model $\mathcal{N}_I(\mathcal{K})$ if and only if x can be represented in the form $x = Az$ for some (generalized block-triangular) matrix $A \in \mathbf{M}^*_{\mathcal{K}}(I)$, where $z = (z_{[K]}|K\in J(\mathcal{K})) \in \mathbb{R}^I$ is an unobservable stochastic variate such that $z \sim N(\Sigma=\text{identity matrix})$. From (4.4), this representation is equivalent to the system of equations

$$(5.13) \qquad x_{[L]} = \Sigma(A_{LM}z_{[M]}|M\in H(L)), \quad L \in J(\mathcal{K}),$$

where $H(L) = \{M\in J(\mathcal{K})|M\leq L\}$. This shows that the CI model $\mathcal{N}_I(\mathcal{K})$ can be interpreted as a multivariate linear recursive model (cf. Wermuth (1980), Kiiveri, Speed, and Carlin (1984)) with lattice constraints.

Conversely, suppose that J is a finite index set and let $(H(\ell)|\ell\in J)$ be a family of subsets of J that satisfies the following two conditions:

(i) $$\ell \in H(\ell)$$

(ii) $$m \in H(\ell) \Rightarrow H(m) \subseteq H(\ell).$$

For each $\ell \in J$ let D_ℓ and E_ℓ be finite index sets such that $|D_\ell| \leq |E_\ell|$ and let $I = \dot{\cup}(D_\ell|\ell\in J)$, $I' = \dot{\cup}(E_\ell|\ell\in J)$. Consider the normal statistical model defined by the system of equations

$$x_{[\ell]} = \Sigma(A_{\ell m} z_{[m]}|\ m\in H(1)),\quad \ell \in J, \tag{5.14}$$

where $x_{[\ell]} \in \mathbf{M}(D_\ell \times\{1\})$ is observable, $z_{[m]} \in \mathbf{M}(E_m\times\{1\})$ is unobservable, $z \equiv (z_{[m]}|m\in J) \sim N(\Sigma=\text{identity matrix})$ on $\mathbb{R}^{I'}$, $A_{\ell m} \in \mathbf{M}(D_\ell\times E_m)$, and $\operatorname{rank}(A_{\ell\ell}) = |D_\ell|$. Let $\mathcal{H}$ be the ring of subsets of J generated by $\{H(\ell)|\ell\in J\}$ and for $H \in \mathcal{H}$ define $I_H = \dot{\cup}(D_\ell|\ell\in H)$; then trivially $\mathcal{R} := \{I_H|H\in\mathcal{H}\}$ is a ring of subsets of I. If we set $\mathcal{K} = \mathcal{K}(\mathcal{R})$ (cf. Example 2.1), then it may be seen that the model determined by the system (5.14) has the form (5.13), i.e., it is the model $\mathcal{N}_I(\mathcal{K})$. □

5.3. Invariance of the model.

It follows from the well-known transformation property of the multivariate normal distribution that the i.i.d. model determined by $\mathcal{N}_I(\mathcal{K})$ is invariant under the transitive action (4.12) of $\mathbf{M}^*_{\mathcal{K}}(I)$ on the parameter space $\mathbf{P}_{\mathcal{K}}(I)$ and the action

$$(5.15) \qquad \begin{aligned} \mathbf{M}^*_{\mathcal{K}}(I) \times \mathbf{M}(I \times N) &\to \mathbf{M}(I \times N) \\ (A, y) &\to Ay \end{aligned}$$

of $\mathbf{M}^*_{\mathcal{K}}(I)$ on the observation (sample) space $\mathbf{M}(I \times N)$. The ML estimator is thus equivariant.

§6. TESTING ONE PAIRWISE CONDITIONAL INDEPENDENCE MODEL AGAINST ANOTHER.

If $\mathcal{K}$ and $\mathcal{M}$ are two distributive sublattices of $\mathcal{K}(V)$ such that $\mathcal{M} \subset \mathcal{K}$, then $\mathbf{P}_{\mathcal{K}}(V) \subseteq \mathbf{P}_{\mathcal{M}}(V)$ and one can consider the statistical problem of testing $\mathcal{N}_V(\mathcal{K})$ against the (possibly) larger model $\mathcal{N}_V(\mathcal{M})$ on the basis of N i.i.d. observations. In this section the central distribution of the likelihood ratio (LR) statistic is derived by means of the invariance of this testing problem under the actions of $GL_{\mathcal{K}}(V)$.

In Section 6.1 the testing problem and the LR statistic Q are given in matrix formulation. The invariance of the testing problem is described in Section 6.2 and used to derive the central moments of Q. This derivation is based on Theorem 6.1, which establishes the mutual independence of the maximal invariant statistic π and the ML estimators $\hat{\Sigma}_{K\bullet}$, $K \in J(\mathcal{K})$. Specific examples of the general testing problem are presented in Section 6.3. Theorem 6.1 is proved in Section 6.4 by means of invariance arguments.

A warning about the notation is needed here. Since $J(\mathcal{K}) \neq J(\mathcal{M})$, quantities such as $J(K)$, D_K, Σ_{KK}, $\Sigma_{K>}$, $\Sigma_{K\bullet}$ depend not only on the quotient space K but also on the lattice of which K is considered a member. Thus, for example, $J_{\mathcal{K}}(K)$ and $J_{\mathcal{M}}(K)$ need not be the same. To alleviate this difficulty without introducing $\mathcal{K}$ and $\mathcal{M}$ as subscripts, the letter K shall denote a quotient that is to be considered as a member of $\mathcal{K}$, while M shall denote a quotient that is to be considered a member of $\mathcal{M}$.

For notational convenience, we sometimes use the following abbreviations: $|K| := |I_K|$ $(K \in \mathcal{K})$, $|K\bullet| := |D_K|$ $(K \in J(\mathcal{K}))$; also $|M| := |I_M|$ $(M \in \mathcal{M})$, $|M\bullet| := |D_M|$ $(M \in J(\mathcal{M}))$.

6.1. Matrix formulation of the testing problem.

It is important to note that if $(e_i|i\in I)$ is a basis for V adapted to $\mathcal{K}$, then it must also be adapted to $\mathcal{M}$ (recall Remark 2.2). To see this, let $\varphi:\mathcal{M} \to \mathcal{K}$ be the embedding of $\mathcal{M}$ into $\mathcal{K}$ and let $\psi = J(\varphi):J(\mathcal{K}) \to J(\mathcal{M})$ be the associated poset homomorphism (cf. [A] (1987), Proposition 1.2 (ii)). Then for $M \in J(\mathcal{M})$ we may define $D_M := \dot{\bigcup}(D_K|K\in J(\mathcal{K}),\ \psi(K)=M)$. Since ψ is surjective ([A] (1987), Proposition 1.3(i)), $I = \dot{\bigcup}(D_M|M\in J(\mathcal{M}))$. Since ψ is a poset homomorphism, it follows that $(e_i|i\in I)$ is also adapted to $\mathcal{M}$.

Once such basis is chosen, the above testing problem may be stated in matrix formulation as follows: based on N i.i.d. observations $x_1,\cdots,x_N \in \mathbb{R}^I$ from the model $\mathcal{N}_I(\mathcal{M})$, test

$$\text{(6.1)} \qquad H_0\colon \Sigma \in \mathbf{P}_{\mathcal{K}}(I) \quad \text{vs.} \quad H\colon \Sigma \in \mathbf{P}_{\mathcal{M}}(I), \qquad (\mathcal{M} \subset \mathcal{K}).$$

The existence of the ML estimator $\tilde{\Sigma}$ $(=: \hat{\Sigma}_{\mathcal{M}})$ under H implies the existence of the ML estimator $\hat{\Sigma}$ $(=: \hat{\Sigma}_{\mathcal{K}})$ under H_0 (recall (5.6)). To see this simply note that

$$\max\{|I_M| \ \ |M\in J(\mathcal{M})\} = \max\{|I_{\psi(K)}| \ \ |K\in J(\mathcal{K})\} \geq \max\{|I_K| \ \ |K\in J(\mathcal{K})\}.$$

The equality follows from the fact that ψ is surjective, while the inequality follows from the relation $\psi(K) \geq K$ for $K \in J(\mathcal{K})$, which is a consequence of the definition of ψ $(= J(\varphi))$ (cf. [A] (1987), equation (1.4)). Thus, from (5.8) and (5.7), the LR statistic Q for testing H_0 against H is given by

$$Q^{2/N} = \frac{\det(\tilde{\Sigma})}{\det(\hat{\Sigma})} \tag{6.2}$$

$$= \frac{\Pi(\det(\tilde{\Sigma}_{M\bullet})\,|M\in J(\mathcal{M}))}{\Pi(\det(\hat{\Sigma}_{K\bullet})\,|K\in J(\mathcal{K}))} = \frac{\Pi(\det(S_{M\bullet})\,|M\in J(\mathcal{M}))}{\Pi(\det(S_{K\bullet})\,|K\in J(\mathcal{K}))}.$$

For computational purposes, note that

$$\det(S_{K\bullet}(y)) = \frac{\det(S_K(y))}{\det(S_{\langle K\rangle}(y))} = \frac{\det(y_K y_K^t)}{\det(y_{\langle K\rangle} y_{\langle K\rangle}^t)}, \tag{6.3}$$

$K \in J(\mathcal{K})$, where $S(y) = yy^t$ and $y = (x_1,\cdots,x_N) \in \mathbf{M}(I\times N)$, with an identical formula for $\det(S_{M\bullet}(y))$, $M \in J(\mathcal{M})$.

6.2. Central distribution of the likelihood ratio statistic.

The testing problem (6.1) is invariant under the action (5.15) of the group $\mathbf{M}_{\mathcal{K}}^{*}(I)$ on the sample space $\mathbf{M}(I\times N)$ and the action

$$\begin{aligned} \mathbf{M}_{\mathcal{K}}^{*}(I)\times\mathbf{P}_{\mathcal{M}}(I) &\to \mathbf{P}_{\mathcal{M}}(I) \\ (A,\Sigma) &\to A\Sigma A^t \end{aligned} \tag{6.4}$$

on the parameter space. Let

$$(6.5) \qquad \pi : \mathbb{M}(I\times N) \to \mathbb{M}(I\times N)/\mathbb{M}^*_{\mathcal{K}}(I)$$

denote the orbit projection ($\equiv$ maximal invariant) onto the orbit space under the action (5.15). Since the LR statistic is invariant under (5.15) Q depends on $y \in \mathbb{M}(I\times N)$ only through $\pi(y)$. The central distribution of Q is readily derived from this fact and Theorem 6.1, whose proof is deferred to Section 6.4. (Since the restriction of (6.4) to $\mathbf{P}_{\mathcal{K}}(I)$ is transitive (cf. (4.12)), under H_0 the distribution of Q does not depen on $\Sigma \in \mathbf{P}_{\mathcal{K}}(I)$.)

Theorem 6.1. Under H_0, the statistics π and $\hat{\Sigma}_{K\bullet}$, $K \in J(\mathcal{K})$, are mutually independent. The statistic $\hat{\Sigma}_{K\bullet}$ has the Wishart distribution on $\mathbf{P}(D_K)$ with $N-|I_{J(K)}|$ degrees of freedom and expected value $\Sigma_{K\bullet}$. □

It follows from Theorem 6.1 that Q and $\hat{\Sigma}_{K\bullet}$, $K \in J(\mathcal{K})$, are mutually independent. Therefore for every $\Sigma \in \mathbf{P}_{\mathcal{K}}(I)$ ($\subseteq \mathbf{P}_{\mathcal{M}}(I)$) and $\alpha \geq 0$,

$$E((\det(\tilde{\Sigma}))^{\alpha}) = E((\det(\hat{\Sigma}))^{\alpha}\, Q^{2\alpha/N}) = E((\det(\hat{\Sigma}))^{\alpha})E(Q^{2\alpha/N}),$$

hence from (4.17) (cf. (6.2)),

$$E(Q^{2\alpha/N}) = \frac{E((\det(\tilde{\Sigma}))^{\alpha})}{E((\det(\hat{\Sigma}))^{\alpha})} = \frac{\Pi(E((\det(\tilde{\Sigma}_{M\bullet}))^{\alpha})\,|M\in J(\mathcal{M}))}{\Pi(E((\det(\hat{\Sigma}_{K\bullet}))^{\alpha})\,|K\in J(\mathcal{K}))}.$$

However, it follows from the Wishart distribution of $\hat{\Sigma}_{K\bullet}$ that

$$E((\det(\hat{\Sigma}_{K\bullet}))^{\alpha}) =$$

$$(2N)^{\alpha|K\bullet|}(\det(\Sigma_{K\bullet}))^{\alpha}\times \Pi\left[\frac{\Gamma((N-|J(K)|-i+1)/2+\alpha)}{\Gamma((N-|J(K)|-i+1)/2)}\middle| i=1,\cdots,|K\bullet|\right]$$

for $K \in J(\mathcal{K})$, with an analogous formula for $E((\det(\tilde{\Sigma}_{M\bullet}))^{\alpha})$, $M \in J(\mathcal{M})$. Since

$$\Sigma(|K\bullet|\,|K\in J(\mathcal{K})) = |I| = \Sigma(|M\bullet|\,|M\in J(\mathcal{M})) \tag{6.6}$$

and

$$\Pi(\det(\Sigma_{K\bullet})\,|K\in J(\mathcal{K})) = \det(\Sigma) = \Pi(\det(\Sigma_{M\bullet})\,|M\in J(\mathcal{M}))$$

for $\Sigma \in P_{\mathcal{K}}(I)$, one obtains that

$$E(Q^{2\alpha/N}) = \frac{\Pi\left[\Pi\left[\frac{\Gamma((N-|J(M)|-i+1)/2+\alpha)}{\Gamma((N-|J(M)|-i+1)/2)}\middle| i=1,\cdots,|M\bullet|\right]\middle| M\in J(\mathcal{M})\right]}{\Pi\left[\Pi\left[\frac{\Gamma((N-|J(K)|-j+1)/2+\alpha)}{\Gamma((N-|J(K)|-j+1)/2)}\middle| j=1,\cdots,|K\bullet|\right]\middle| K\in J(\mathcal{K})\right]}. \tag{6.7}$$

The Box approximation for the central distribution of $-2\log Q$ may be obtained as in Anderson (1984) p.311-316. In Anderson's notation we have $a = b = |I|$ and

$$(6.8)\qquad \begin{aligned} f &= -2\Sigma(\Sigma((-|J(M)|-i+1)/2\,|\,i=1,\cdots,|M\cdot|)\,|\,M\in J(\mathcal{M})) \\ &\quad +2\Sigma(\Sigma((-|J(K)|-j+1)/2\,|\,j=1,\cdots,|K\cdot|)\,|\,K\in J(\mathcal{K})) \\ &= \Sigma(|M\cdot|\times|J(M)|+|M\cdot|(|M\cdot|-1)/2\,|\,M\in J(\mathcal{M})) \\ &\quad -\Sigma(|K\cdot|\times|J(K)|+|K\cdot|(|K\cdot|-1)/2\,|\,K\in J(\mathcal{K})) \\ &= \Sigma(|M\cdot|\times|J(M)|+|M\cdot|(|M\cdot|+1)/2\,|\,M\in J(\mathcal{M})) \\ &\quad -\Sigma(|K\cdot|\times|J(K)|+|K\cdot|(|K\cdot|+1)/2\,|\,K\in J(\mathcal{K})) \end{aligned}$$

where the final equality is obtained using (6.6). From (3.13), one recognizes f to be simply the usual difference between the number of free parameters under H and the number of free parameters under H_0.

6.3. Examples of testing problems.

Let $\mathcal{K}_1,\cdots,\mathcal{K}_8,\mathcal{K}_9,\mathcal{K}_{10},\mathcal{K}_{11}$ denote the lattices appearing in Figures 3.1,•••,3.8,3.9a,3.10a,3.11a, respectively. In this subsection we shall consider examples of the testing problem (6.1) with $(\mathcal{K},\mathcal{M}) = (\mathcal{K}_i,\mathcal{K}_j)$ for various pairs (i,j). In each example the LR statistic Q in (6.2) and the parameter f in (6.8) will be rewritten in forms that reflect the statistical interpretation of the testing problem, i.e., that reflect the conditional independence (CI) condition being tested.

For this purpose we must introduce the following notation: for any $\Sigma \in P(I)$ and any $K,L \in \mathcal{K}$ such that $L \leq K$, let

$$\Sigma_K = \begin{bmatrix} \Sigma_L & \Sigma_{L,K\setminus L} \\ \Sigma_{K\setminus L,L} & \Sigma_{K\setminus L} \end{bmatrix}$$

denote the partitioning of Σ_K according to the decomposition

$$I_K = I_L \dot{\cup} (I_K \setminus I_L)$$

and define

$$\Sigma_{K\cdot L} = \Sigma_{K\setminus L} - \Sigma_{K\setminus L,L}\Sigma_L^{-1}\Sigma_{L,K\setminus L}.$$

(When $K \in J(\mathcal{K})$ and $M \in J(\mathcal{M})$, $\Sigma_{K\cdot J(K)} = \Sigma_{K\cdot}$ and $\Sigma_{M\cdot J(M)} = \Sigma_{M\cdot}$.) The well-known formula

$$\det(S_{K\cdot L}) = \det(S_K)/\det(S_L)$$

may be applied in (6.2) to obtain the expressions for $Q^{2/N}$ that appear below.

First, consider the testing problems of the form

(6.9) $$H_0\colon \Sigma \in P_{\mathcal{K}}(I) \quad \text{vs.} \quad H\colon \Sigma \in P(I)$$

for $\mathcal{K} = \mathcal{K}_3, \cdots, \mathcal{K}_8$. (Note that $P_{\mathcal{K}}(I) = P(I)$ for $\mathcal{K} = \mathcal{K}_1$ and $\mathcal{K} = \mathcal{K}_2$.) In each of these problems the following form of the LR statistic Q directly reflects the statistical interpretation of the model $N_I(\mathcal{K})$ given in Section 5.2.:

$\mathcal{K} = \mathcal{K}_3$:

$$Q_3^{2/N} = \frac{\det(S)}{\det(S_L)\det(S_M)},$$

$$f_3 = |I_L| \times |I_M|;$$

$\mathcal{H} = \mathcal{H}_4$: $$Q_4^{2/N} = \frac{\det(S_{LVM})}{\det(S_L)\det(S_M)},$$

$$f_4 = |I_L| \times |I_M|;$$

$\mathcal{H} = \mathcal{H}_5$: $$Q_5^{2/N} = \frac{\det(S_{V\cdot(L\wedge M)})}{\det(S_{L\cdot(L\wedge M)})\det(S_{M\cdot(L\wedge M)})},$$

$$f_5 = |D_L| \times |D_M|;$$

$\mathcal{H} = \mathcal{H}_6$: $$Q_6^{2/N} = \frac{\det(S_{(LVM)\cdot(L\wedge M)})}{\det(S_{L\cdot(L\wedge M)})\det(S_{M\cdot(L\wedge M)})},$$

$$f_5 = |D_L| \times |D_M|;$$

$\mathcal{H} = \mathcal{H}_7$: $$Q_7^{2/N} =$$

$$\frac{\det(S_{(LVM)\cdot(L\wedge M)})}{\det(S_{L\cdot(L\wedge M)})\det(S_{M\cdot(L\wedge M)})} \times \frac{\det(S_{V\cdot(LVM)})}{\det(S_{L'\cdot(LVM)})\det(S_{M'\cdot(LVM)})}$$

$$f_7 = |D_L| \times |D_M| + |D_{L'}| \times |D_{M'}|;$$

$\mathcal{H} = \mathcal{H}_8$: $$Q_8^{2/N} = \frac{\det(S_{(LVM)\cdot(L\wedge M)})}{\det(S_{L\cdot(L\wedge M)})\det(S_{M\cdot(L\wedge M)})}$$

$$\times \frac{\det(S_{L'\cdot L})}{\det(S_{(LVM)\cdot L})\det(S_{L''\cdot L})} \times \frac{\det(S_{V\cdot(LVM)})}{\det(S_{L'\cdot(LVM)})\det(S_{M'\cdot(LVM)})}$$

$$= \frac{\det(S_{(LVM)\cdot(L\wedge M)})}{\det(S_{L\cdot(L\wedge M)})\det(S_{M\cdot(L\wedge M)})} \times \frac{\det(S_{V\cdot L})}{\det(S_{L''\cdot L})\det(S_{M'\cdot L})}$$

$$= \frac{\det(S_{L'\cdot(L\wedge M)})}{\det(S_{M\cdot(L\wedge M)})\det(S_{L''\cdot(L\wedge M)})} \times \frac{\det(S_{V\cdot(L\vee M)})}{\det(S_{L'\cdot(L\vee M)})\det(S_{M'\cdot(L\vee M)})}$$

$$\begin{aligned} f_8 &= |D_L|\times|D_M| + |D_M|\times|D_{L''}| + |D_{L''}|\times|D_{M'}| \\ &= |D_L|\times|D_M| + |D_{L''}|(|D_M|+|D_{M'}|) \\ &= |D_M|(|D_L|+|D_{L''}|) + |D_{L''}|\times|D_{M'}|; \end{aligned}$$

Remark 6.1. The three equivalent expressions for $Q_8^{2/N}$ given above correspond to the first three minimal determining sets of CI conditions for $\mathcal{N}_I(\mathcal{K})$ given in Remark 5.1. The expression for $Q_8^{2/N}$ suggested by the fourth set is

$$\frac{\det(S_{L'\cdot(L\wedge M)})}{\det(S_{M\cdot(L\wedge M)})\det(S_{L''\cdot(L\wedge M)})} \times \frac{\det(S_{V\cdot L})}{\det(S_{L''\cdot L})\det(S_{M'\cdot L})}$$

but this is not equal to $Q_8^{2/N}$. Thus the fourth determining set is in some sense unsatisfactory for describing $\mathcal{N}_I(\mathcal{K})$. □

Next we consider five testing problems of the form (6.1) with $(\mathcal{K},\mathcal{M}) = (\mathcal{K}_i,\mathcal{K}_j)$. From (6.2) and (6.8) one may obtain the following expressions:

$$(\mathcal{K},\mathcal{M}) = (\mathcal{K}_7,\mathcal{K}_6): \qquad Q_{7,6}^{2/N} = (Q_7/Q_6)^{2/N} = \frac{\det(S_{V\cdot(L\vee M)})}{\det(S_{L'\cdot(L\vee M)})\det(S_{M'\cdot(L\vee M)})},$$

$$f_{7,6} = f_7 - f_6 = |D_{L'}|\times D_{M'}|;$$

$$(\mathcal{H},\mathcal{M}) = (\mathcal{H}_8,\mathcal{H}_7): \qquad Q_{8,7}^{2/N} = (Q_8/Q_7)^{2/N} = \frac{\det(S_{L'\cdot L})}{\det(S_{(LVM)\cdot L})\det(S_{L''\cdot L})},$$

$$f_{8,7} = f_8 - f_7 = |D_M|\times|D_{L''}|;$$

$$(\mathcal{H},\mathcal{M}) = (\mathcal{H}_9,\mathcal{H}_8): \qquad Q_{9,8}^{2/N} = \frac{\det(S_{M'\cdot M})}{\det(S_{(LVM)\cdot M})\det(S_{M''\cdot M})},$$

$$f_{9,8} = |D_L|\times D_{M''}|;$$

$$(\mathcal{H},\mathcal{M}) = (\mathcal{H}_{11},\mathcal{H}_6): \quad Q_{11,6}^{2/N} = \frac{\det(S_{L'\cdot L})}{\det(S_{(LVM)\cdot L})\det(S_{L''\cdot L})},$$

$$f_{11,8} = |D_M|\times|D_{L''}|;$$

$$(\mathcal{H},\mathcal{M}) = (\mathcal{H}_8,\mathcal{H}_{11}): \quad Q_{8,11}^{2/N} = Q_{7,6}^{2/N}, \qquad f_{8,11} = f_{7,6};$$

These five testing problems involve the five adjacent pairs of lattices in the diagram

$$\mathcal{H}_6 \begin{matrix} \subset & \mathcal{H}_7 & \subset \\ \subset & \mathcal{H}_{11} & \subset \end{matrix} \mathcal{H}_8 \subset \mathcal{H}_9.$$

The LR statistic Q and the parameter f for non-adjacent pairs may be obtained from those for adjacent pairs in the usual way, for example:

$$(\mathcal{H},\mathcal{M}) = (\mathcal{H}_9,\mathcal{H}_7): \qquad Q_{9,7}^{2/N} = (Q_{9,8}\cdot Q_{8,7})^{2/N}, \qquad f_{9,7} = f_{9,8} + f_{8,7}.$$

Remark 6.2. It is thus seen that in each example, the LR statistic can be represented as a product of LR statistics for testing CI of two blocks of variates. We conjecture that this is true in general, i.e., that the LR statistic Q in (6.2) for the general testing problem (6.1) may be written as such a product, and that furthermore, the factors are mutually independent under H_0. Of course it must be realized that the above examples involve only very simple lattices. More complex distributive lattices, e.g. non-planar lattices, may lead to statistical models and tests with more complex structure. □

Remark 6.3. Each of the testing problems treated by Das Gupta (1977), Banerjee and Giri (1980), and Marden (1981) is a special case of the general testing problems (6.1) or (6.9). □

6.4. Proof of Theorem 6.1.

Let $\Omega \subset \mathbf{M}(I\times N)$ be the open subset

$$\Omega = \{y\in\mathbf{M}(I\times N) \mid \operatorname{rank}(y) = \min\{|I|,N\}\}. \tag{6.10}$$

Since $\mathbf{M}(I\times N)\setminus\Omega$ is a Lebesgue-null set, we may replace the sample space $\mathbf{M}(I\times N)$ by Ω. Also, since $\operatorname{rank}(Ay) = \operatorname{rank}(y)$ for $A \in \mathbf{M}_{\mathcal{K}}^{*}(I)$ and $y \in \mathbf{M}(I\times N)$, it follows that $\mathbf{M}_{\mathcal{K}}^{*}(I)$ acts on Ω by restriction of (5.15). Furthermore, since Ω is locally compact, Lemma 6.2 at the end of this subsection implies that this restriction is a proper action (whereas (5.15) itself is not proper). Thus, in order to prove Theorem 6.1 we may apply the method of [A] (1982) to study the transformation of the normal distributions in the model H_0 under the mapping

$$(6.11) \qquad \begin{aligned} \Omega &\to \Omega/\mathrm{M}^{*}_{\mathcal{K}}(I) \times (\mathrm{X}(P(D_K)) \,|\, K\in J(\mathcal{K}))) \\ y &\to (\pi(y) \;,\; (\hat{\Sigma}_{K\bullet}(y) \,|\, K\in J(\mathcal{K}))). \end{aligned}$$

The group $\mathrm{M}^{*}_{\mathcal{K}}(I)$ is the semidirect product of its two closed subgroups $\mathcal{A}$ and $\mathcal{T}$, where

$$\begin{aligned} \mathcal{A} &= \{A \in \mathrm{M}^{*}_{\mathcal{K}}(I) \,|\, A_{K>} = 0, \quad K \in J(\mathcal{K})\} \\ \mathcal{T} &= \{T \in \mathrm{M}^{*}_{\mathcal{K}}(I) \,|\, T_{KK} = 1_{KK}, \; K \in J(\mathcal{K})\}. \end{aligned}$$

Therefore we may apply the method of [A] (1982), Section 5, with $K = \mathrm{M}^{*}_{\mathcal{K}}(I)$, $H = \mathcal{A}$, $G = \mathcal{T}$, and $X = \Omega$ to see that π can be represented as $\pi = \pi_{\mathcal{A}} \circ \pi_{\mathcal{T}}$, where $\pi_{\mathcal{T}}:\Omega \to \Omega/\mathcal{T}$ and $\pi_{\mathcal{A}}:\Omega/\mathcal{T} \to (\Omega/\mathcal{T})/\mathcal{A} \simeq \Omega/\mathrm{M}^{*}_{\mathcal{K}}(I)$. (The action of $\mathcal{T}$ on Ω is the restriction of (5.15) to $\mathcal{T}\times\Omega$, and the induced action of $\mathcal{A}$ on $\Omega/\mathcal{T}$ is defined as in equation (21) of [A] (1982).)

Since the mapping (6.11) is invariant under the action of $\mathcal{T}$ on Ω (cf. (4.16)), it has a unique factorization through $\pi_{\mathcal{T}}$. Therefore we may first transform the normal distributions in the model H_0 from Ω to $\Omega/\mathcal{T}$ by $\pi_{\mathcal{T}}$. To do this, we need the following explicit representation:

<u>Lemma 6.1.</u> A representation of $\pi_{\mathcal{T}}:\Omega \to \Omega/\mathcal{T}$ is given by

$$(6.12) \quad \Omega/\mathcal{T} = \{y=(y_{[K]} \,|\, K\in J(\mathcal{K})) \in \Omega \,|\, y_{[K]}y^{t}_{\langle K\rangle}(y_{\langle K\rangle}y^{t}_{\langle K\rangle})^{-1}y_{\langle K\rangle} = 0, \; K\in J(\mathcal{K})\},$$

$$(6.13) \qquad \pi_{\mathcal{T}}(y) = (y_{[K]} - y_{[K]}y^{t}_{\langle K\rangle}(y_{\langle K\rangle}y^{t}_{\langle K\rangle})^{-1}y_{\langle K\rangle} \,|\, K\in J(\mathcal{K})).$$

Proof: To show that $\Omega/\mathcal{T}$ in (6.12) is a cross-section of Ω and that $\pi_{\mathcal{T}}$ in (6.13) is a maximal invariant function, it suffices to show that for each $y \in \Omega$,

$$\{Ty \mid T\in\mathcal{T}\} \cap (\Omega/\mathcal{T}) = \{\pi_{\mathcal{T}}(y)\}. \tag{6.14}$$

To show the inclusion $\subseteq$ in (6.14), suppose that $Ty \in \Omega/\mathcal{T}$. Then from (6.12), (4.7), and (4.8),

$$\begin{aligned} 0 &= (Ty)_{[K]}(Ty)^t_{\langle K\rangle}((Ty)_{\langle K\rangle}(Ty)^t_{\langle K\rangle})^{-1}(Ty)_{\langle K\rangle} \\ &= (y_{[K]} + T_{K\rangle}y_{\langle K\rangle})y^t_{\langle K\rangle}T^t_{\langle K\rangle}(T_{\langle K\rangle}y_{\langle K\rangle}y^t_{\langle K\rangle}T^t_{\langle K\rangle})^{-1}T_{\langle K\rangle}y_{\langle K\rangle} \\ &= (y_{[K]} + T_{K\rangle}y_{\langle K\rangle})y^t_{\langle K\rangle}(y_{\langle K\rangle}y^t_{\langle K\rangle})^{-1}y_{\langle K\rangle} \end{aligned}$$

for each $K \in J(\mathcal{K})$, hence

$$(Ty)_{[K]} = y_{[K]} - y_{[K]}y^t_{\langle K\rangle}(y_{\langle K\rangle}y^t_{\langle K\rangle})^{-1}y_{\langle K\rangle}, \tag{6.15}$$

$K \in J(\mathcal{K})$, i.e., $Ty = \pi_{\mathcal{T}}(y)$. To show the opposite inclusion $\supseteq$, it is easy to verify that $\pi_{\mathcal{T}}(y) \in (\Omega/\mathcal{T})$ for every $y \in \Omega$; to show that $\pi_{\mathcal{T}}(y) \in \{Ty \mid T\in\mathcal{T}\}$, simply note that $\pi_{\mathcal{T}}(y) = \tilde{T}y$ where

$$\tilde{T}_{K\rangle} = -y_{[K]}y^t_{\langle K\rangle}(y_{\langle K\rangle}y^t_{\langle K\rangle})^{-1},$$

$K \in J(\mathcal{K})$. Finally, the mapping $\pi_{\mathcal{T}}:\Omega \to \Omega/\mathcal{T}$ defined in (6.12) and (6.13) is clearly continuous, so this representation is also topological and the result follows. □

We may now apply formula (16) of [A] (1982) to transform the normal distributions in the model H_0 by the mapping $\pi_{\mathcal{T}}$ (6.13). In the notation of [A] (1982), $G = \mathcal{T}$, $X = \Omega$, λ is the restriction of Lebesgue measure on $\mathbf{M}(I\times N)$ to the open subset Ω, $\pi = \pi_{\mathcal{T}}$, β is a Haar measure on $\mathcal{T}$, $\Delta_G = \Delta_{\mathcal{T}} \equiv 1$, and $P = p\cdot\lambda$, where p is the density given by

$$p(y) = (\det(\Sigma))^{-N/2}\exp\{-\mathrm{tr}(\Sigma^{-1}yy^t)/2\},\ y \in \Omega.$$

For $\Sigma \in P_{\mathcal{K}}(I)$, the density q of $\pi_{\mathcal{T}}(P)$ with respect to the quotient measure λ/β on $\Omega/\mathcal{T}$ is thus given by

$$(6.16)\quad q(\pi_{\mathcal{T}}(y)) = (\det(\Sigma))^{-N/2} \int_{\mathcal{T}}\exp\{-\mathrm{tr}(\Sigma^{-1}(Ty)(Ty)^t)/2\}d\beta(T)$$

$$= \Pi((\det(\Sigma_{K\cdot}))^{-N/2}\exp\{-\mathrm{tr}(\Sigma_{K\cdot}^{-1}y_{K\cdot}y_{K\cdot}^t)/2\} \mid K\in J(\mathcal{K}))$$

$$\times\int_{\mathcal{T}}\Pi(\exp\{-\mathrm{tr}(\Sigma_{K\cdot}^{-1}z_K(T)z_K(T)^t)/2\} \mid K\in J(\mathcal{K}))d\beta(T),$$

where

$$y_{K\cdot} = (y_{[K]} - y_{[K]}y_{\langle K\rangle}^t(y_{\langle K\rangle}y_{\langle K\rangle}^t)^{-1}y_{\langle K\rangle}),$$

$$z_K(T) = (y_{[K]} + T_{K\rangle}y_{\langle K\rangle})y_{\langle K\rangle}^t(y_{\langle K\rangle}y_{\langle K\rangle}^t)^{-1}y_{\langle K\rangle}-\Sigma_{K\rangle}\Sigma_{\langle K\rangle}T_{\langle K\rangle}y_{\langle K\rangle},$$

$K \in J(\mathcal{K})$, $T \in \mathcal{T}$.

Since $d\beta(T) = \Pi(d\lambda_K(T_{K>})|K\in J(\mathcal{K}))$, where λ_K is the Lebesgue measure on $M(D_K \times I_{J(K)})$ (cf. (4.9), the last integral in (6.16) can be calculated using Fubini's Theorem and the translation invariance of λ_K, $K \in J(\mathcal{K})$. The order of integration should be determined by a never-increasing listing $K_1, K_2, \cdots, K_{|J(\mathcal{K})|}$ of the elements in $J(\mathcal{K})$ (cf. Remark 3.6). After some calculation we obtain

$$(6.17)\quad q(\pi_{\mathcal{T}}(y)) = \Pi((\det(\Sigma_{K\cdot}))^{-N/2}\exp\{-tr(\Sigma_{K\cdot}^{-1}y_{K\cdot}y_{K\cdot}^{t})/2\}\mid K\in J(\mathcal{K}))$$

$$\times\Pi((\det(\Sigma_{K\cdot}))^{|J(K)|/2}(\det(y_{\langle K\rangle}y_{\langle K\rangle}^{t}))^{-|K\cdot|/2}\mid K\in J(\mathcal{K}))$$

$$= \Pi((\det(\Sigma_{K\cdot}))^{-(N-|J(K)|)/2}\exp\{-tr(\Sigma_{K\cdot}^{-1}S_{K\cdot}(y))/2\}\mid K\in J(\mathcal{K}))$$

$$\times\Pi((\det S_{\langle K\rangle}(y))^{-|K\cdot|/2}\mid K\in J(\mathcal{K})),\qquad \pi_{\mathcal{T}}(y) \in \Omega/\mathcal{T},$$

where $S(y) = yy^t$.

By Lemma 6.1 wherein $\Omega/\mathcal{T}$ is representated as a subset of Ω, the induced action of the subgroup $\mathcal{A}$ on $\Omega/\mathcal{T}$ is simply the restriction of the action (5.15) to $\mathcal{A}\times(\Omega/\mathcal{T})$. The next step is to represent the transformed measure $\pi_{\mathcal{T}}(P) = q\cdot(\lambda/\beta)$ as $\pi_{\mathcal{T}}(P) = q_1\cdot\nu$, where ν is an invariant measure under this action of $\mathcal{A}$ on $\Omega/\mathcal{T}$.

It follows from a statement on p. 961 of [A] (1982) that the quotient measure λ/β is relatively invariant under the action of $\mathcal{A}$ on $\Omega/\mathcal{T}$ with multiplier χ given by $\chi(A) = (\mathrm{mod}\varphi_A)^{-1}\chi_0(A)$, $A \in \mathcal{A}$, where χ_0 is the multiplier for λ as a relatively invariant measure under the action of $M_{\mathcal{K}}^{*}(I)$ on Ω and where the automorphisms $\varphi_A:\mathcal{T} \to \mathcal{T}$ are defined by $\varphi_A(T) = ATA^{-1}$, $T \in \mathcal{T}$. Since $A = \mathrm{Diag}(A_{KK}|K\in J(\mathcal{K}))$ it is clear that

$$(\varphi_A(T))_{K\rangle} = A_{KK}T_{K\rangle}A^{-1}_{\langle K\rangle},$$

$K \in J(\mathscr{K})$, hence

$$\mathrm{mod}\varphi_A = \Pi(|\det(A_{KK})|^{|J(K)|}|\det(A_{\langle K\rangle})|^{|K\bullet|}|\ K\in J(\mathscr{K})).$$

But also

$$\chi_0(A) = |\det(A)|^N = \Pi(|\det(A_{KK})|^N|\ K\in J(\mathscr{K})),$$

so that

$$\chi(A) = \Pi(|\det(A_{KK})|^{N-|J(K)|}|\det(A_{\langle K\rangle})|^{|K\bullet|}|\ K\in J(\mathscr{K})).$$

If we define $n:\Omega/\mathscr{T}\to]0,\infty[$ by

$$n(\pi_{\mathscr{T}}(y)) = \Pi((\det(\hat{\Sigma}_{K\bullet}(y)))^{(N-|J(K)|)/2}(\det(S_{\langle K\rangle}(y)))^{|K\bullet|/2}|\ K\in J(\mathscr{K})),$$

it follows that $n(Az) = \chi(A)n(z)$, $z \in \Omega/\mathscr{T}$, $A \in \mathscr{A}$. Thus the measure $\nu :=$ $n^{-1}\bullet(\lambda/\beta)$ is invariant under the action of $\mathscr{A}$ on $\Omega/\mathscr{T}$. From (6.17), the density $q_1 = nq$ of $\pi_{\mathscr{T}}(P)$ with respect to ν is therefore given by

$$q_1(\pi_{\mathscr{T}}(x)) =$$

$$\Pi\left(\left[\frac{\det(\hat{\Sigma}_{K\bullet}(y))}{\det(\Sigma_{K\bullet})}\right]^{(N-|J(K)|)/2} \times \exp\{-N\mathrm{tr}(\Sigma^{-1}_{K\bullet}\hat{\Sigma}_{K\bullet}(y))/2\}\ \middle|\ K\in J(\mathscr{K})\right),$$

where it should be recalled that $\Sigma \in P_{\mathscr{K}}(I)$.

The final step in the proof of Theorem 6.1 is to obtain the transformation of the measure $\pi_{\mathscr{T}}(P) = q_1\bullet\nu$ under the mapping

$$(6.18) \qquad \begin{aligned} \Omega/\mathcal{T} &\to (\Omega/\mathcal{T})/\mathcal{A} \times (X(P(D_K)\,|K\in J(\mathcal{K}))) \\ \pi_{\mathcal{T}}(y) &\to (\pi_{\mathcal{A}}(\pi_{\mathcal{T}}(y)), (\hat{\Sigma}_{K\bullet}(y)\,|K\in J(\mathcal{K}))). \end{aligned}$$

Since the action of $\mathcal{A}$ on $\Omega/\mathcal{T}$ is the restriction to the closed subset $\mathcal{A}\times(\Omega/\mathcal{T})$ of the proper action of $M^{*}_{\mathcal{K}}(I)$ on Ω, it is a proper action. Thus we may apply Lemma 3 of Andersson, Brøns and Jensen (1983) to see that there exists a unique measure κ on $(\Omega/\mathcal{T})/\mathcal{A}$ such that the invariant measure ν is transformed into the product measure $\kappa\otimes\nu_0$ under the mapping (6.18), where ν_0 is an invariant measure on $X(P(D_K)\,|K\in J(\mathcal{K}))$ under the proper and transitive action

$$(6.19) \qquad \begin{aligned} \mathcal{A}\times(X(P(D_K)\,|K\in J(\mathcal{K}))) &\to X(P(D_K)\,|K\in J(\mathcal{K})) \\ (A\ ,\ (\Lambda_K\,|K\in J(\mathcal{K}))) &\to (A_{KK}\Lambda_K A^t_{KK}\,|K\in J(\mathcal{K})). \end{aligned}$$

(Lemma 3 of Andersson, Brøns, and Jensen (1983) is applied with $G = \mathcal{A}$, $X = \Omega/\mathcal{T}$, $Y = X(P(D_K)\,|K\in J(\mathcal{K}))$, $t = (\pi_{\mathcal{T}}(y) \to (\hat{\Sigma}_{K\bullet}(y)\,|K\in J(\mathcal{K})))$, $\pi = \pi_{\mathcal{A}}$, and $\nu = \nu$.)

Since $q_1(z)$ depends on $z := \pi_{\mathcal{T}}(y)$ only through $(\hat{\Sigma}_{K\bullet}(y)\,|K\in J(\mathcal{K}))$, the probability measure $q_1\bullet\nu$ is therefore transformed under (6.18) into the probability measure $r\bullet(\kappa\otimes\nu_0)$, where

$$\begin{aligned} r:(\Omega/\mathcal{T})/\mathcal{A} \times (X(P(D_K)\,|K\in J(\mathcal{K}))) &\to \mathbb{R}_+ \\ (w, (\Lambda_K\,|K\in J(\mathcal{K}))) &\to \end{aligned}$$

$$(6.20) \qquad \Pi\left(\left[\frac{\det(\Lambda_K)}{\det(\Sigma_{K\bullet})}\right]^{(N-|J(K)|)/2} \times \exp\{-N\mathrm{tr}(\Sigma^{-1}_{K\bullet}\Lambda_K)/2\}\,|K\in J(\mathcal{K})\right).$$

Because r does not depend on w, under H_0 it follows that $\pi = \pi_{\mathscr{A}} \circ \pi_{\mathscr{T}}$ is independent of $(\hat{\Sigma}_{K\cdot} | K \in J(\mathscr{K}))$, π has distribution κ, and $(\hat{\Sigma}_{K\cdot} | K \in J(\mathscr{K}))$ has distribution $s \cdot \nu_0$, where $s((\Lambda_K | K \in J(\mathscr{K}))$ is given by the product (6.20). Furthermore, since $\nu_0 = \otimes(\nu_K | K \in J(\mathscr{K}))$ where ν_K is an invariant measure on $P(D_K)$ under the usual action of $M^*(D_K)$, it follows that under H_0, $\hat{\Sigma}_{K\cdot}$, $K \in J(\mathscr{K})$, are mutually independent and $\hat{\Sigma}_{K\cdot}$ has the Wishart distribution on $P(D_K)$ with $N-|J(K)|$ degrees of freedom and expected value $\Sigma_{K\cdot}$. This concludes the proof of Theorem 6.1.

The following lemma, which was cited at the beginning of this subsection, is also of interest in its own right for the study of group actions in statistics. (see also Bourbaki (1971), Chapitre III, §4, Proposition 5 (ii)).

Lemma 6.2. Suppose that G and G' are locally compact groups that act continuously on the locally compact spaces X and X', respectively. Let $\varphi: G \to G'$ be a continuous group homomorphism and $\psi: X \to X'$ be a continuous mapping such that $\psi(gx) = \varphi(g)\psi(x)$, $x \in X$, $g \in G$. If φ is proper and if the action of G' on X' is proper, then the action of G on X is also proper.

Proof: Consider the diagram

$$\begin{array}{ccc} G \times X & \xrightarrow{\vartheta} & X \times X \\ {\scriptstyle \varphi\times\psi} \downarrow & & \downarrow {\scriptstyle \psi\times\psi} \\ G' \times X' & \xrightarrow{\vartheta'} & X' \times X', \end{array}$$

where $\vartheta(g,x)=(gx,x)$ and $\vartheta'(g',x')=(g'x',x')$. We must show that $\vartheta^{-1}(C)$ is compact whenever $C \subseteq X\times X$ is compact. Let $p_{G'}$ denote the projection of $G'\times X'$ onto G'. Since the diagram commutes, i.e., $\vartheta'\circ(\varphi\times\psi)=(\psi\times\psi)\circ\vartheta$, it follows that

$$\begin{aligned}\vartheta^{-1}(C) &\subseteq \vartheta^{-1}((\psi\times\psi)^{-1}((\psi\times\psi)(C))) = (\varphi\times\psi)^{-1}(\vartheta'^{-1}((\psi\times\psi)(C)))\\ &\subseteq (\varphi\times\psi)^{-1}(p_{G'}(\vartheta'^{-1}((\psi\times\psi)(C)))\times X) = \varphi^{-1}(C')\times X\end{aligned}$$

where $C' = p_{G'}(\vartheta'^{-1}((\psi\times\psi)(C)))$. Since trivially $\vartheta^{-1}(C) \subseteq G\times p_2(C)$, where p_2 denotes the projection of $X\times X$ on the second component, we have that

$$\vartheta^{-1}(C) \subseteq \varphi^{-1}(C')\times p_2(C).$$

But C' is compact since ϑ' is proper and therefore $\varphi^{-1}(C')$ is compact because φ is proper. Thus $\vartheta^{-1}(C)$ is a closed subset of a compact subset of $G\times X$, hence is compact. □

With the identifications $G = G' = \mathbf{M}^{*}_{\mathscr{K}}(I)$, $X = \Omega$, $X' = \mathbf{P}_{\mathscr{K}}(I)$, $\varphi =$ the identity mapping on $\mathbf{M}^{*}_{\mathscr{K}}(I)$, and $\psi = \hat{\Sigma}$, Lemma 6.2 may be applied as indicated at the beginning of this subsection.

§7. CONCLUDING REMARKS.

A more detailed investigation of the structure of the normal conditional independence (CI) models $\mathscr{N}_V(\mathscr{K})$ and the associated testing problems will be presented in a subsequent study. Among the questions under investigation is that of characterizing the minimal determining sets of CI conditions for $\mathcal{N}_V(\mathcal{K})$ (cf. Remark 5.1). A second question is whether every testing problem of the general form (6.1) can be decomposed into a product of simpler testing problems (cf. Remark 6.2). The answer to this question will be of use for a decision-theoretic study of the LR test and other invariant tests for the problem (6.1).

The normal statistical models $\mathcal{N}_V(\mathcal{K})$ may be generalized in several ways. One natural and possibly fruitful extension is suggested by an examination of the $\mathcal{K}$-parametrization (3.13) of $P_{\mathcal{K}}(I)$ once a $\mathcal{K}$-adapted basis for V has been chosen. A large class of "second-order" submodels of $\mathcal{N}_I(\mathcal{K})$ may be obtained by replacing each $P(D_K)$ in (3.13) by $P_{\mathcal{K}'}(D_K)$, where each $\mathcal{K}' = \mathcal{K}_K$ is a distributive sublattice of $\mathcal{K}(\mathbb{R}^{D}K)$. Third-order and higher-order submodels may be obtained by iterating this process. This construction yields a very rich and varied class of normal conditional models and associated testing problems which, despite their apparent complexity, admit a relatively standard explicit likelihood analysis.

Alternatively, one might replace each term $M(D_K \times I_{J(K)}) \times P(D_K)$ in the $\mathcal{K}$-parametrization (3.13) by a suitable covariance selection model requirement (cf. Dempster (1972), Wermuth (1976, 1980)), thus generalizing the multivariate graphical chain models of Lauritzen and Wermuth (1987) to "multivariate graphical lattice models".

Another question currently under investigation is the relation of the lattice models $\mathcal{N}_V(\mathcal{K})$ and their extensions to the normal models for CI determined by recursive causal graphs and decomposable graphs (cf. Wermuth (1980, 1985), Kiiveri, Speed, and Carlin (1984), Lauriten (1985, 1989), Lauritzen and Wermuth (1987)).

REFERENCES.

[1] Andersen, A.H. (1974). Multidimensional contingency tables. Scand. J. Statist. 1, 115-127.

[2] Andersen, E.B. (1980). Discrete Statistical Models with Social Science Applications. North Holland, Amsterdam.

[3] Anderson, T.W. (1985). An Introduction to Multivariate Statistical Analysis (2nd ed.). Wiley, New York.

[4] Andersson, S.A. (1982). Distributions of maximal invariants using quotient measures. Ann. Statist. 10, 955-961.

[5] Andersson, S.A. (1987). The lattice structure of orthogonal linear models and orthogonal variance component models. Tech. Report No. 123, Dept. of Statistics, University of Washington, Seattle, Washington.

[6] Andersson, S.A., Brøns, H.K., and Jensen, S.T. (1983). Distribution of eigenvalues in multivariate statistical analysis. Ann. Statist. 11, 392-415.

[7] Andersson, S.A., Marden, J.I., and Perlman, M.D. (1988). Totally ordered multivariate linear models with applications to nested missing data problems. In preparation.

[8] Andersson, S.A., Marden, J.I., and Perlman, M.D. (1989). Partially ordered multivariate linear models with applications to non-nested missing data problems. In preparation.

[9] Banerjee, P.K., and Giri, N. (1980). On D-, E-, D_A-, and D_M-optimality properties of test procedures of hypotheses concerning the covariance matrix of a normal distribution. In Multivariate Statistical Analysis (R.P. Gupta, ed.) 11-19, North Holland Pub. Co., New York.

[10] Bourbaki, N. (1971). Éléments de Mathématique. Topologie generale. Chap. 1 á 4. Herman, Paris.

[11] Darroch, J.N., Lauritzen, S.L., and Speed, T.P. (1980). Markov fields and log-linear interaction models for contingency tables. Ann. Statist. 8, 522-539.

[12] Darroch, J.N. and Speed, T.P. (1983). Additive and multiplicative models and interactions. Ann. Statist. 11, 724-738.

[13] Das Gupta, S. (1977). Tests on multiple correlation coefficient and multiple partial correlation coefficient. J. Multivariate Analysis 7, 82-88.

[14] Dawid, A.P. (1979). Conditional independence in statistical theory (with discussion). J. Roy. Statist. Soc. Ser. B 41, 1-41.

[15] Dawid, A.P. (1980). Conditional independence for statistical operations. Ann. Statist. 8, 598-617.

[16] Dempster, A. (1972). Covariance selection models. Biometrics 28, 157-175.

[17] Eaton, M.L. (1983). Multivariate Statistics: A Vector Space Approach. Wiley, New York.

[18] Eaton, M.L. and Kariya, T. (1983). Multivariate tests with incomplete data. Ann. Statist. 11, 654-665.

[19] Goodman, L.A. (1970). The multivariate analysis of qualitative data: interaction among multiple classifications. J. Amer. Statist. Assoc. 65, 226-256.

[20] Goodman, L.A. (1971). Partitioning of chi-square, analysis of marginal contingency tables, and estimation of expected frequencies in multidimensional contingency tables. J. Amer. Statist. Assoc. 66, 339-344.

[21] Haberman, S. (1974). The Analysis of Frequency Data. University of Chicago Press, Chicago, Illinois.

[22] Kiiveri, H., Speed, T.P., and Carlin, J.B. (1984). Recursive causal models. J. Austral. Math. Soc. (Ser. A) 36, 30-52.

[23] Lauritzen, S.L. (1985). Test of hypotheses in decomposable mixed interaction models. Bull. Int. Statist. Inst. 4, 24.3(1)-24.3(6).

[24] Lauritzen, S.L. (1989). Mixed graphical association models. To appear in Scand. J. Statist.

[25] Lauritzen, S.L., Dawid, A.P., Larsen, B.N. and Leimer, H.G. (1988). Independence Properties of Directed Markov Fields. Res. rep. R-88-32. Inst. of Electr. Systems. Aalborg Univ.

[26] Lauritzen, S.L. and Frydenberg, M. (1988). Decomposition of Maximum Likelihood in Mixed Graphical Interaction Models. Res. rep. R-88-17. Inst. of Electr. Systems. Aalborg Univ.

[27] Lauritzen, S.L. and Wermuth, N. (1984). Mixed interaction models. Res. rep. R-84-8. Inst. of Electr. Systems. Aalborg Univ.

[28] Lauritzen, S.L. and Wermuth, N. (1987). Graphical models for association between variables, some of which are qualitative and some quantitative. Res. rep. R-87-10. Inst. of Electr. Systems. Aalborg Univ. To appear in Ann. Statist.

[29] Marden, J.I. (1981). Invariant tests on covariance matrices. Ann. Statist. 9, 1258-1266.

[30] Speed, T.P. and Kiiveri, H. (1986). Gaussian Markov distributions over finite graphs. Ann. Statist. 16, 138-150.

[31] Sundberg, R. (1975). Some results about decomposable (or Markov-type) models for multidimensional contingency tables: distribution of marginals and partitioning of tests. Scand. J. Statist. 2, 71-79.

[32] Wermuth, N. (1976). Analogies between multiplicative models in contingency tables and covariance selection. Biometrics 32, 95-108.

[33] Wermuth, N. (1980). Linear recursive equations, covariance selection, and path analysis. J. Amer. Statist. Assoc. 75, 963-972.

[34] Wermuth, N. (1985). Data analysis and conditional independence structures. Bull. Int. Statist. Inst. 4, 24.2(1)-24.2(13).

[35] Wermuth, N. (1988). On Block-Recursive Linear Regression Equations. Manuscript. Psychologisches Institut. Universität Mainz.

[36] Wermuth, N. and Lauritzen, S.L. (1983). Graphical and recursive models for contingency tables. Biometrika 70, 537-552.

PREPRINTS 1987

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK, TELEPHONE + 45 1 35 31 33.

No. 1 Jensen, Søren Tolver and Johansen, Søren: Estimation of Proportional Covariances.

No. 2 Rootzén, Holger: Extremes, Loads, and Strengths.

No. 3 Bertelsen, Aksel: On the Problem of Testing Reality of a Complex Multivariate Normal Distribution.

No. 4 Gill, Richard D. and Johansen, Søren: Product-Integrals and Counting Processes.

No. 5 Leadbetter, M.R. and Rootzén, Holger: Extremal Theory for Stochastic Processes.

No. 6 Tjur, Tue: Block Designs and Electrical Networks.

No. 7 Johansen, Søren: Statistical Analysis of Cointegration Vectors.

No. 8 Bertelsen, Aksel: On the Problem of Testing Reality of a Complex Multivariate Normal Distribution, II.

No. 9 Andersson, S.A. and Perlman, M.D.: Group-Invariant Analogues of Hadamard's Inequality.

No. 10 Hald, Anders: Two Generalizations of the Problem of Points by Bernoulli, de Moivre and Montmort.

No. 11 Andersson, Steen Arne: The Lattice Structure of Orthogonal Linear Models and Orthogonal Variance Component Models.

PREPRINTS 1988

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK, TELEPHONE + 45 1 35 31 33.

No. 1 Jacobsen, Martin: Discrete Exponential Families: Deciding when the Maximum Likelihood Estimator Exists and Is Unique.

No. 2 Johansen, Søren and Juselius, Katarina: Hypothesis Testing for Cointegration Vectors - with an Application to the Demand for Money in Denmark and Finland.

No. 3 Jensen, Søren Tolver, Johansen, Søren and Lauritzen, Steffen L.: An Algorithm for Maximizing a Likelihood Function.

No. 4 Bertelsen, Aksel: On Non-Null Distributions Connected with Testing that a Real Normal Distribution Is Complex.

No. 5 Tjur, Tue: Statistical Tables for Personal Computer Users.

No. 6 Tjur, Tue: A New Upper Bound for the Efficiency of a Block Design.

No. 7 Bunzel, Henning, Høst, Viggo and Johansen, Søren: Some Simple Non-Parametric Tests for Misspecification of Regression Models Using Sign Changes of Residuals.

No. 8 Brøns, Hans and Jensen, Søren Tolver: Maximum Likelihood Estimation in the Negative Binomial Distribution.

No. 9 Andersson, S.A. and Perlman, M.D.: Lattice Models for Conditional Independence in a Multivariate Normal Distribution.