LATTICE MODELS FOR CONDITIONAL INDEPENDENCE IN A MULTIVARIATE NORMAL DISTRIBUTION

by

Steen Arne Andersson
Michael D. Perlman

TECHNICAL REPORT No. 155 (Revised)
August 1991

Department of Statistics, GN-22
University of Washington
Seattle, Washington 98195 USA
LATTICE MODELS FOR CONDITIONAL INDEPENDENCE IN A MULTIVARIATE NORMAL DISTRIBUTION$^{1,2}$

BY

STEEN ARNE ANDERSSON$^3$

DEPARTMENT OF MATHEMATICS

UNIVERSITY OF INDIANA

AND

MICHAEL D. PERLMAN

DEPARTMENT OF STATISTICS

UNIVERSITY OF WASHINGTON

---

$^1$This research was supported in part by the Danish Research Council and by U.S. National Science Foundation Grant Nos. DMS 86-03489 and 89-02211.

$^2$Revised August 1991.

$^3$This author's research was carried out at the Institute of Mathematical Statistics, University of Copenhagen.
Summary

The lattice conditional independence model $\mathcal{N}(\mathcal{A})$ is defined to be the set of all normal distributions on $\mathbb{R}^I$ such that for every pair $L, M \in \mathcal{A}$, $x_L$ and $x_M$ are conditionally independent given $x_{L \cap M}$. Here $\mathcal{A}$ is a lattice of subsets of the finite index set $I$ and, for $K \in \mathcal{A}$, $x_K$ is the coordinate projection of $x \in \mathbb{R}^I$ to $\mathbb{R}^K$. Statistical properties of $\mathcal{N}(\mathcal{A})$ are studied, e.g., maximum likelihood inference, invariance, and the problem of testing $H_0: \mathcal{N}(\mathcal{A})$ vs $H: \mathcal{N}(\mathcal{M})$ when $\mathcal{M}$ is a sublattice of $\mathcal{A}$. The set $J(\mathcal{A})$ of join-irreducible elements of $\mathcal{A}$ plays a central role in the analysis of $\mathcal{N}(\mathcal{A})$. This class of statistical models is relevant to the analysis of non-nested multivariate missing data patterns.


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

§1. INTRODUCTION .................................................................1

§2. THE CLASS $P_\mathcal{X}(I)$ OF COVARIANCE MATRICES $\Sigma$ DETERMINED BY PAIRWISE
CONDITIONAL INDEPENDENCE WITH RESPECT TO A FINITE DISTRIBUTIVE LATTICE $\mathcal{X}$ 13
2.1. The poset $J(\mathcal{X})$ of join-irreducible elements ......................14
2.2. The $\mathcal{X}$-parameters of $\Sigma$ .....................................16
2.3. Characterization of conditional independence in terms of $\Sigma^{-1}$ ....17
2.4. The $\mathcal{X}$-preserving matrices: generalized block-triangular
matrices with lattice structure ..............................................19
2.5. The $\mathcal{X}$-parametrization of $P_\mathcal{X}(I)$ ..........................24
2.6. Transitive action of the group of $\mathcal{X}$-preserving matrices ..........26
2.7. Reconstruction of $\Sigma$ from its $\mathcal{X}$-parameters ......................28
2.8. Examples .................................................................32

§3. LIKELIHOOD INFERENCE FOR A NORMAL MODEL DETERMINED BY PAIRWISE
CONDITIONAL INDEPENDENCE ..............................................47
3.1. Factorization of the likelihood function; the MLE of $\Sigma$ ............47
3.2. Examples of pairwise conditional independence models ................49
3.3. Invariance of the model ................................................56

§4. TESTING ONE PAIRWISE CONDITIONAL INDEPENDENCE MODEL AGAINST ANOTHER 56
4.1. The likelihood ratio statistic .........................................57
4.2. Central distribution and Box approximation ..........................58
4.3. Examples of testing problems .........................................61

§5. INARIANT FORMULATION OF THE CI MODEL AND TESTING PROBLEM 65
5.1. The lattice structure of quotient spaces ..............................65
5.2. Invariant formulation of the pairwise CI model ......................67
5.3. Reduction of the CI model to canonical coordinate-wise form ........68
5.4. Invariant formulation of the testing problem .........................69

§6. CONCLUDING REMARKS .....................................................69

APPENDIX .................................................................71
A.1. The Decomposition Theorem and existence of a $\mathcal{X}$-adapted basis 71
A.2. Proof of Theorem 5.1 ..................................................75
A.3. Proof of Theorem 4.2 ..................................................76

REFERENCES .................................................................85
§1. INTRODUCTION.

Because conditional independence (CI) plays an increasingly important role in statistical model building, it is of interest to study classes of CI models with tractable statistical properties and to develop methods for testing one CI model against another. In this paper we define and study a class of CI models determined by finite distributive lattices.

For multivariate normal distributions, the parameter space and the likelihood function (LF) for such a lattice CI model can be factored into a product of parameter spaces and conditional LF's, respectively, corresponding to ordinary multivariate normal linear regression models. This in turn yields explicit maximum likelihood estimators (MLE) and likelihood ratio tests (LRT) by means of standard technique from multivariate analysis.

These lattice CI models arise in a natural way in the analysis of multivariate missing data sets with non-monotone missing data patterns. The factorizations mentioned above can be readily applied to obtain explicit MLE's and LRT's by standard linear methods (cf. [AP] (1991)1).

We introduce this class of lattice CI models by means of the following simple and familiar model. Let \((x_1, x_2, x_3)^t\) denote a random observation from the trivariate normal distribution \(N(\Sigma)\) with mean vector 0 and unknown covariance matrix \(\Sigma\).2 Consider the model that specifies that \(x_2\)
and $x_3$ are conditionally independent given $x_1$, which we express in the familiar notation

\[(1.1) \quad x_2 \parallel x_3 \mid x_1.\]

In terms of the covariance matrix $\Sigma$, (1.1) is equivalent to the condition

\[(1.2) \quad (\Sigma^{-1})_{23} = (\Sigma^{-1})_{32} = 0.\]

In order to express this as a lattice CI model, let $I \equiv \{1,2,3\}$ denote the index set and consider

\[(1.3) \quad \mathcal{A} \equiv \{\emptyset, \{1\}, \{1,2\}, \{1,3\}, I\},\]

a subring of the ring $\mathcal{B}(I)$ of all subsets of $I$. Clearly $\mathcal{A}$ is a finite distributive lattice under the usual set operations $\cup$ and $\cap$. Define the class $P_{\mathcal{A}}(I)$ of real positive definite $I \times I$ matrices as follows:

\[(1.4) \quad \Sigma \in P_{\mathcal{A}}(I) \iff x_L \parallel x_L \mid x_{L \cup M} \quad \forall L, M \in \mathcal{A},\]

where $x \sim N(\Sigma)$ and $x_T$ denotes the $T$-subvector of $x$ when $T \subseteq I$. It is readily verified that (1.1), (1.2), and (1.4) are equivalent conditions.\(^3\)

In this example the factorizations of the parameter space and LF mentioned above are represented as follows:

\[^3\text{Note that } x_2 \parallel x_3 \mid x_1 \iff (x_1,x_2) \parallel (x_1,x_3) \mid x_1.\]
\[(1.5) \quad \Sigma \leftrightarrow (\Sigma_{11}, \Sigma_{21}\Sigma_{11}^{-1}, \Sigma_{22}\cdot 1, \Sigma_{31}\Sigma_{11}^{-1}, \Sigma_{33}\cdot 1)\]

\[(1.6) \quad f(x_1, x_2, x_3) = f(x_1)f(x_2|x_1)f(x_3|x_1).\]

The five parameters on the right-hand side of (1.5) represent ordinary unconditional and conditional variances and regression coefficients. Whereas the range of the positive definite matrix \(\Sigma\) in (1.5) is constrained by (1.2), the ranges of these five parameters are unconstrained (except for the trivial requirement that \(\Sigma_{11}, \Sigma_{22}\cdot 1,\) and \(\Sigma_{33}\cdot 1\) are positive). Thus the MLE's of these five parameters, called the \(\mathcal{X}\)-parameters of the CI model, are easily obtained from (1.6), and the MLE of \(\Sigma\) may be reconstructed from these estimates.

A subset \(K \in \mathcal{X}\) is called join-irreducible if \(K\) is not the join (union) of two or more proper subsets of \(K\) (cf. Section 2.1). The collection of all join-irreducible elements in \(\mathcal{X}\) is denoted by \(J(\mathcal{X})\). Thus when \(\mathcal{X}\) is given by (1.3),

\[(1.7) \quad J(\mathcal{X}) = \{\{1\}, \{1, 2\}, \{1, 3\}\}.\]

It will be seen that the basic factorizations (1.5) and (1.6), as well as their extensions to the general lattice CI model \(\mathcal{M}(\mathcal{X})\) defined next, always are indexed by the members of \(J(\mathcal{X})\).

Condition (1.4) immediately extends to define the general lattice CI model. Let \(I\) be an arbitrary finite index set and let \(\mathcal{X}\) be an arbitrary subring of \(\mathcal{D}(I)\), so again \(\mathcal{X}\) is a finite distributive lattice. Then (1.4) defines the class \(P_\mathcal{X}(I)\) of \(I \times I\) covariance matrices determined by CI...

---

\[\text{By Birkhoff's Theorem, any finite distributive lattice can be represented as a ring of subsets of some finite set } I.\]
restrictions with respect to the lattice \( \mathcal{X} \): \( \Sigma \in P_{\mathcal{X}}(I) \) if and only if \( x_L \) and \( x_M \) are conditionally independent given \( x_{L \cap M} \) for every pair \( L, M \in \mathcal{X} \).

If \( N(\Sigma) \) denotes the normal distribution on \( \mathbb{R}^I \) with mean vector 0 and unknown covariance matrix \( \Sigma \), the normal statistical model

\[
(1.8) \quad N(\mathcal{X}) \equiv \{N(\Sigma) | \Sigma \in P_{\mathcal{X}}(I)\}
\]

is the lattice conditional independence (CI) model determined by \( \mathcal{X} \).

In this paper we study the structure of \( P_{\mathcal{X}}(I) \) and the statistical properties of the model \( N(\mathcal{X}) \). In Section 2.3 (Theorem 2.1) we generalize (1.2) by characterizing \( \Sigma \in P_{\mathcal{X}}(I) \) in terms of the precision matrix \( \Sigma^{-1} \).

In Section 2.5 (Theorem 2.2) we generalize (1.5) by showing that each \( \Sigma \in P_{\mathcal{X}}(I) \) can be uniquely represented in terms of its \( \mathcal{X} \)-parameters, whose range are unconstrained, so that the parameter space \( P_{\mathcal{X}}(I) \) again factors into a product of parameter spaces for ordinary linear regression models.

In Section 2.7 we present a general algorithm for reconstructing \( \Sigma \in P_{\mathcal{X}}(I) \) from its \( \mathcal{X} \)-parameters. A series of examples in Section 2.8 illustrates these results.

The factorization (1.6) of the LF as a product of conditional densities involving only the \( \mathcal{X} \)-parameters of \( \Sigma \) is extended to the general lattice CI model \( N(\mathcal{X}) \) in Section 3.1 (Theorem 3.1). The MLE's of the \( \mathcal{X} \)-parameters of \( \Sigma \) are easily derived from the general factorization, then the MLE of \( \Sigma \) can be reconstructed by the algorithm given in Section 2.7. This estimation procedure is illustrated by examples in Section 3.2. In Remark 3.5 it is noted that the model \( N(\mathcal{X}) \) is determined by a system of linear recursive equations (cf. Wermuth (1980) or Kiiveri, Speed, and Carlin (1984)) with additional lattice structure.
In Section 4 we treat the problem of testing one lattice CI model against another, i.e., testing

\[(1.9) \quad H_0: \mathcal{N}(\mathcal{M}) \quad \text{vs.} \quad H: \mathcal{N}(\mathcal{M})\]

when \(\mathcal{M}\) is a sublattice of \(\mathcal{X}\).\(^5\) For example, in the trivial case considered above with \(I = \{1,2,3\}\), suppose that \(\mathcal{X} = \{\emptyset, \{1\}, \{1,2\}, \{1,3\}, I\}\) (cf. (1.3)) and \(\mathcal{M} = \{\emptyset, I\}\). Then \(\mathcal{N}(\mathcal{M})\) is simply the normal model with no restriction on \(\Sigma\) and (1.9) becomes the problem of testing \(x_2 \parallel x_3 \mid x_1\) (equivalently, (1.2)) against the unrestricted alternative, which can be stated equivalently as the problem of testing

\[(1.10) \quad H_0: \sigma_{23} = \sigma_{21}^{-1}\sigma_{11}\sigma_{13} \quad \text{vs.} \quad H: \sigma_{23} \neq \sigma_{21}^{-1}\sigma_{11}\sigma_{12}.\]

where \(\Sigma = (\sigma_{ij} \mid i,j = 1,2,3)\). If, however,

\[(1.11) \quad \mathcal{X} = \{\emptyset, \{1\}, \{3\}, \{1,2\}, \{1,3\}, I\}\]

while \(\mathcal{M} = \{\emptyset, \{1\}, \{1,2\}, \{1,3\}, I\}\), then (1.9) becomes the problem of testing \((x_1,x_2) \parallel x_3\) against \(x_2 \parallel x_3 \mid x_1\), which is equivalent to the problem of testing

---

\(^5\)Note that \(\mathcal{M} \subseteq \mathcal{X} \Rightarrow \mathcal{N}(\mathcal{X}) \subseteq \mathcal{N}(\mathcal{M})\).

\(^6\)When \(\mathcal{X}\) is given by (1.11) and \(\mathcal{X}' = \{\emptyset, \{3\}, \{1,2\}, I\}\) then \(\mathcal{N}(\mathcal{X}) = \mathcal{N}(\mathcal{X}')\): both lattices determine the same CI conditions, namely \((x_1,x_2) \parallel x_3\). Thus two different lattices may determine the same CI model.
The LRT statistic $\lambda$ for the general testing problem (1.9) is derived in Section 4.1 and is readily expressible in terms of the MLE's of the $\Psi$-parameters and $\Psi$-parameters of $\Sigma$. In Section 4.2 the central distribution of $\lambda$ is derived in terms of its moments by means of the invariance of the testing problem. Specific examples of this testing problem are considered in Section 4.3.

These and associated results are greatly facilitated by the fact that the model $\mathcal{M}(\Psi)$ is invariant under a group $G \equiv \text{GL}_{\Psi}(I)$ that $G$ acts transitively on $P_{\Psi}(I)$. This group $G$ is a subgroup of a group of nonsingular block-triangular $I \times I$ matrices. To illustrate this, return to the trivariate lattice CI model considered above with $\Psi$ given by (1.3). It can be seen that the CI model given by (1.1) $\equiv$ (1.2) is invariant under all nonsingular linear transformations of the form

\begin{equation}
(1.13) \quad 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 A x.
\end{equation}

and that any nonsingular linear transformation $A$ that leaves this CI model invariant must be of the form (1.13). The collection of all such matrices $A$ forms a subgroup of the group of all $3 \times 3$ nonsingular lower triangular matrices. It is also true, but not so easy to see, that $G$ acts

\footnote{In fact, each of the testing problems treated by Das Gupta (1977), Giri (1979), Banerjee and Giri (1980), and Marden (1981), including (1.12), is a special case of the general problem (1.9).}
transitively on the class \( P_\mathfrak{M}(I) \) of all covariance matrices \( \Sigma \) that satisfy
\((1.1) \equiv (1.2)\), i.e., for any such \( \Sigma \) there exists \( A \in G \) such that \( \Sigma = AA^t \).

These facts, some of which were used by Das Gupta (1977), Giri (1979), Banerjee and Giri (1980), and Marden (1981) to study the distribution and optimality of invariant tests for problems such as (1.10) and (1.12), will be extended in the present paper to the general lattice CI model \( \mathcal{N}(\mathfrak{M}) \). In Section 2.4 it will be shown how \( \mathfrak{M} \) determines the invariance group \( \text{GL}_\mathfrak{M}(I) \), a group of generalized block-triangular \( I \times I \) matrices with lattice structure, while the transitive action of \( \text{GL}_\mathfrak{M}(I) \) on \( P_\mathfrak{M}(I) \) is demonstrated in Section 2.6 (Theorem 2.3), generalizing the well-known Choleski decomposition of an arbitrary positive definite matrix. The transitivity yields a factorization (Lemma 2.5) of the determinant of \( \Sigma \in P_\mathfrak{M}(I) \), a generalization of the well-known Schur formula \( \det(\Sigma) = \det(\Sigma_{11})\det(\Sigma_{22,1}) \).

As already seen for the trivariate example above, all statistical properties of the general lattice CI model \( \mathcal{N}(\mathfrak{M}) \), including the definition of the \( \mathfrak{M} \)-parameters of \( \Sigma \), the factorizations of its parameter space and LF as products of those for linear regression models, the form of the MLE, the form of the LRT statistic and its central distribution, and the partitioning and location of zeroes in the invariance matrix \( A \in \text{GL}_\mathfrak{M}(I) \), are determined by the fundamental structure of the lattice \( \mathfrak{M} \), in particular by the associated poset \( J(\mathfrak{M}) \) of join-irreducible elements of \( \mathfrak{M} \) (cf. Section 2.1). As in the case of a balanced ANOVA design where the poset of join-irreducible elements of the lattice of subspaces determines the ANOVA table (cf. [A] (1990)), for the lattice CI model \( \mathcal{N}(\mathfrak{M}) \) the poset \( J(\mathfrak{M}) \) determines the statistical analysis of the model.

The CI models \( \mathcal{N}(\mathfrak{M}) \) play an important role in the analysis of
non-monotone missing data models. Under the assumption of multivariate normality it is well known that a monotone missing data model with unrestricted covariance matrix $\Sigma$ admits a complete and explicit likelihood analysis, remaining invariant under the appropriate group of block-triangular matrices (in the usual sense), which acts transitively on the unrestricted set of covariance matrices (cf. Eaton and Kariya (1983), [AMP] (1990)). If the missing data pattern is non-monotone, however, then explicit analysis is not possible in general.

The relationship between lattice CI models and non-monotone missing data patterns is developed fully in [AP] (1991) but can be illustrated in terms of the trivariate example considered above. Suppose that one attempts to observe a random sample from the trivariate normal distribution $N(\Sigma)$, where $\Sigma$ is unknown and initially unrestricted, but that some of the observations are incomplete. For example, suppose that we have several complete vector observations of the form $(x_1, x_2, x_3)^t$ and also several incomplete observations of the forms $(x_1, x_2)^t$ and $(x_1, x_3)^t$. Then the missing data pattern (actually, the pattern of the observed data) is the set

$\mathcal{Y} := \{\{1,2\}, \{1,3\}, \{1,2,3\}\}$.

(1.14)

i.e., the collection of subsets of $I \equiv \{1,2,3\}$ corresponding to the subvectors actually observed. Because the missing data pattern $\mathcal{Y}$ is non-monotone, i.e., is not totally ordered by inclusion, the LF cannot be factored into a product of LF's of linear regression models and the MLE
of $\Sigma$ cannot be obtained explicitly. Instead, iterative estimation methods such as the EM algorithm must be used, possibly accompanied by difficulties with convergence or uniqueness of the estimates (cf. Little and Rubin (1987)).

An alternate approach, suggested by Rubin (1987) and developed in [AP] (1991), is to restrict $\Sigma$ by imposing the CI conditions of the lattice CI model $N(\mathfrak{X})$, where $\mathfrak{X} \equiv \mathfrak{X}(\mathfrak{Y})$ is the lattice generated by $\mathfrak{Y}$. With $\mathfrak{Y}$ given by (1.14) it is easy to see that $\mathfrak{X}$ is given by (1.3), so the corresponding CI condition is given by (1.1). Under this condition the densities for the complete and incomplete observations factor as

$$f(x_1, x_2, x_3) = f(x_1)f(x_2|x_1)f(x_3|x_1),$$

(1.15) $$f(x_1, x_2) = f(x_1)f(x_2|x_1),$$

$$f(x_1, x_3) = f(x_1)f(x_3|x_1),$$

so the overall LF is a product of LF's of only the three types $f(x_1)$, $f(x_2|x_1)$, and $f(x_3|x_1)$, the latter two corresponding to simple linear regression models. Also, the overall parameter space is the product of the parameter spaces for these three LF's. Therefore the similar terms may be combined and the MLE of $\Sigma$ may be obtained by maximizing these three LF's separately, which involves only elementary calculations.

Furthermore, under the CI restriction $\Sigma \in P_{\mathfrak{X}}(I)$, this non-monotone missing data model remains invariant under the group $GL_{\mathfrak{X}}(I)$ of lower triangular matrices $A$ in (1.13) and $GL_{\mathfrak{X}}(I)$ acts transitively on $P_{\mathfrak{X}}(I)$.

---

8In fact, for some non-monotone missing data patterns with insufficiently many complete observations, $\Sigma$ may not be estimable.
Finally, the CI assumption may be tested by means of the LRT for (1.10) as discussed above.

Whereas the determination of the appropriate CI conditions and the factorization (1.15) is transparent in this simple example, a general missing data pattern requires the lattice-theoretic approach developed in the present paper — see [AP] (1991) for complete details. 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 monotone class.

In Section 5 the CI models and results already described are recast in an invariant (≡ coordinate-free) formulation, rather than in the matrix (coordinate-wise) formulation just given. This is done for the following reason: a model which, when presented in matrix formulation, may not appear to be a lattice CI model according to the non-invariant definition given above, may in fact belong to this class after an appropriate linear transformation.\(^9\)

This is readily illustrated in terms of the trivariate missing data example given in the paragraph containing (1.14). Rather than the missing data pattern described by (1.14), consider a missing data array that includes incomplete observations involving not only the coordinates of \(x\) but also one or more linear combinations of these coordinates. For example, suppose that we have several complete observations of the form \((x_1,x_2,x_3)^t\) and also several incomplete observations of the forms

\(^9\)Of course this is by no means unique to the lattice CI models. For example, the general balanced ANOVA model must be described invariantly in terms of a collection of orthogonal subspaces, rather than by specifying the values of certain coordinates of the mean vector.
Although this does not directly fit into the framework of the coordinate-wise missing data models discussed above and in [AP] (1991), it is easy to transform it to such a framework by means of a nonsingular linear transformation \((y_1, y_2, y_3) = (x_1 + x_2, x_2, x_3)\). In terms of \(y_1, y_2, y_3\) the missing data pattern is now given precisely by (1.14), hence as before the associated lattice CI model imposes the assumption that \(y_2 \parallel y_3 \mid y_1\), i.e., \(x_2 \parallel x_3 \mid x_1 + x_2\) (equivalently, \(x_1 \parallel x_3 \mid x_1 + x_2\)).

The existence and form of an appropriate linear transformation from \(x\) to \(y\) (or equivalently, of an appropriate vector basis for the observation space) may not be so apparent in more complex missing data schemes with linear combinations present. The invariant formulation of a general lattice CI model, presented in Section 5, allows one to recognize and treat, without a preliminary transformation, a set of CI conditions such as \(x_2 \parallel x_3 \mid x_1 + x_2\) in the same manner as the coordinate-wise lattice CI conditions in (1.4).

The invariant formulation is stated in terms of a lattice \(Q\) of quotient spaces \(Q\) of a real finite-dimensional vector space \(V\). For each \(Q \in \mathcal{Q}\) let \(p_Q : V \to Q\) denote the projection onto \(Q\). Then the general lattice CI model \(\mathcal{N}_V(Q)\) is defined in Section 5.2 to be the set of all nonsingular normal distributions \(N(\sigma)\) on \(V\) with mean \(0\) and covariance \(\sigma\) such that \(p_R\) and \(p_T\) are conditionally independent given \(p_{R\cap T}\) for every pair \(R, T \in \mathcal{Q}\). In Theorem 5.1 it is noted that \(\mathcal{N}_V(Q)\) is nonempty if and only if \(Q\) is distributive.

---

10 All vector spaces and matrices considered in this paper are defined over the field of real numbers.
To express our original coordinate-wise formulation of the lattice CI models in this invariant framework, set \( V = \mathbb{R}^I \), identify each subset \( K \subseteq I \) with the quotient space \( \mathbb{R}^K \), and let \( p^K_K : \mathbb{R}^I \to \mathbb{R}^K \) denote the usual coordinate projection mapping. Then the definition of the general lattice CI model in the preceding paragraph reduces to (1.4).

The basic decomposition theorem for a distributive lattice \( Q \) of quotient spaces (cf. Appendix A.1) states that the observation space \( V \) can be represented as a product of vector spaces indexed by the poset \( J(Q) \) of join-irreducible elements in \( Q \) in such a way that for each \( Q \in Q \), the projection \( p_Q : V \to Q \) becomes simply a canonical projection. By means of this representation we may choose a \( Q \)-adapted basis for \( V \) (cf. Proposition 5.1). In Section 5.3 it is shown that in terms of the coordinate system determined by this basis, the CI model \( \mathcal{N}_V(Q) \) can be expressed in the canonical coordinate-wise form (1.4) and the statistical analysis of the model may then proceed according to the coordinate-wise formulation.

The general problem of testing one lattice CI model against another is formulated invariantly as follows: test \( H_0 : \mathcal{N}_V(Q) \) vs. \( H : \mathcal{N}_V(\mathcal{F}) \), where \( Q \) and \( \mathcal{F} \) are distributive lattices of quotient spaces of \( V \) such that \( \mathcal{F} \subseteq Q \). In Section 5.4 it is noted that one can choose a basis for \( V \) that is both \( Q \)-adapted and \( \mathcal{F} \)-adapted, by means of which this testing problem can be reduced to the canonical coordinate-wise form (1.9).

Several possible extensions of the class of lattice CI models are discussed briefly in Section 6. Three important but technical theorems are proved in the Appendix.

In recent years the study of multivariate dependence models defined by CI conditions determined by directed or undirected graphs has received
increasing attention. Prominent references for normal distributions include Dempster (1972), Frydenberg (1990), Frydenberg and Lauritzen (1989), Kiiveri, Speed, and Carlin (1984), Lauritzen (1985, 1989), Lauritzen, Dawid, Larsen and Leimer (1990), Lauritzen and Wermuth (1984, 1989), Porteous (1985), Speed and Kiiveri (1986), and Wermuth (1976, 1980, 1985, 1988); see Whittaker (1990) for a readable introduction to this area. In many of these studies the CI assumptions are equivalent to the occurrence of patterns of zeroes in the precision matrix $\Sigma^{-1}$ of a multivariate normal distribution, hence the models are linear in $\Sigma^{-1}$. It will be seen from Examples 2.6 - 2.8, however, that unlike the special case (1.2), in general the lattice CI models introduced here are neither linear in $\Sigma^{-1}$ nor $\Sigma$. Furthermore, the statistical interpretation and analysis of a lattice CI model appear to differ from those of a model defined by graphical conditions. Although it is of interest to determine the relation between these two types of CI models and compare their properties, our attempts to interpret either class in the framework of the other have not been illuminating thus far.

§2. THE CLASS $P_{\mathcal{M}}(I)$ OF COVARIANCE MATRICES $\Sigma$ DETERMINED BY PAIRWISE CONDITIONAL INDEPENDENCE WITH RESPECT TO A FINITE DISTRIBUTIVE LATTICE $\mathcal{M}$.

Let $I$ be a finite index set, let $\mathcal{Z}(I)$ denote the ring of all subsets of $I$, and let $\mathcal{M} \subseteq \mathcal{Z}(I)$ be a subring, i.e., $\mathcal{M}$ is closed under $\cap$ and $\cup$. We shall always assume that $I, \emptyset \in \mathcal{M}$. Then $\mathcal{M}$ is a finite distributive lattice with $\cup$ and $\cap$ as the join and meet operations, respectively. For $T, U \in \mathcal{Z}(I)$ we write $T \subseteq U$ to indicate that $T \subseteq U$ but $T \neq U$. Let $|T|$ denote the number of elements in a set $T$. 
Let \( N(\Sigma) \) denote the normal distribution on \( \mathbb{R}^I \) with mean \( 0 \in \mathbb{R}^I \) and covariance matrix \( \Sigma \in \mathcal{P}(I) \), where \( \mathcal{P}(I) \) denotes the set of all positive definite \( I \times I \) matrices. For any \( T \subseteq I \) and column vector \( x = (x_i | i \in I) \in \mathbb{R}^I \) define \( x_T := (x_i | i \in T) \), the \( T \)-subcolumn of \( x \). Note that \( x_I = x \) and define \( x_{\emptyset} := \{0\} \).

**Definition 2.1.** The class \( \mathcal{P}_{\mathcal{X}}(I) \subseteq \mathcal{P}(I) \) is defined as follows (cf. (1.4)):

\[
\Sigma \in \mathcal{P}_{\mathcal{X}}(I) \iff x_L \perp x_M | x_{L \cap M} \quad \forall \, L, M \in \mathcal{X} \text{ when } x \sim N(\Sigma),
\]

i.e., \( x_L \) and \( x_M \) are conditionally independent (CI) given \( x_{L \cap M} \) \( \forall \, L, M \in \mathcal{X} \).

If \( L \cap M = \emptyset \) then (2.1) reduces to \( x_L \perp x_M \) that is, \( x_L \) and \( x_M \) are independent. Note that the right hand side of (2.1) is ordinarily written in the form

\[
x_{L \setminus (L \cap M)} \perp x_{M \setminus (L \cap M)} | x_{L \cap M} \quad \forall \, L, M \in \mathcal{X},
\]

Some of these pairwise CI conditions are trivially satisfied, e.g., whenever \( L \subseteq M \) (or \( M \subseteq L \)) (also see Remark 3.2). In particular, if \( \mathcal{X} \) is a chain then \( \mathcal{P}_{\mathcal{X}}(I) = \mathcal{P}(I) \), i.e., \( \Sigma \) is unrestricted (cf. Examples 2.1 and 2.2).

2.1. The poset \( J(\mathcal{X}) \) of join-irreducible elements.

The structure of \( \Sigma \in \mathcal{P}_{\mathcal{X}}(I) \) will be characterized of terms of the poset \( J(\mathcal{X}) \) of join-irreducible elements of \( \mathcal{X} \), which we now define. For \( K \in \mathcal{X}, K \neq \emptyset \), define
\[
\langle K \rangle := \bigcup \{ K' \in \mathcal{F} | K' \subseteq K \} \\
[K] := K \setminus \langle K \rangle ,
\]
so that
\[
(2.3) \quad K = \langle K \rangle \cup [K].
\]
where \( \cup \) indicates that the union is disjoint. Then define
\[
J(\mathcal{F}) := \{ K \in \mathcal{F} | K \neq \emptyset, \langle K \rangle \subseteq K \}
\]
\[
= \{ K \in \mathcal{F} | K \neq \emptyset, [K] \neq \emptyset \}
\]
\[
= \{ K \in \mathcal{F} | K \neq \emptyset, \forall L, M \in \mathcal{F}: K = L \cup M \Rightarrow K = L \text{ or } K = M \}. 
\]
If \( K \in J(\mathcal{F}) \) we say that \( K \) is join-irreducible. (See Grätzer (1978), Chapter II, or Davey and Priestley (1990), Chapter 8, for properties of \( J(\mathcal{F}) \); in particular, \( \mathcal{F} \) is uniquely determined by \( J(\mathcal{F}) \).)

For \( L \in \mathcal{F} \) define \( \mathcal{F}_L := \{ K \in \mathcal{F} | K \subseteq L \} \), a sublattice of \( \mathcal{F} (\mathcal{F}_L \equiv \mathcal{F}) \). The following relations are elementary:

\[
(2.4) \quad L = \bigcup \{ K \in J(\mathcal{F}_L) \}
\]
\[
(2.5) \quad J(\mathcal{F}_L) = J(\mathcal{F}) \cap \mathcal{F}_L
\]
\[
(2.6) \quad J(\mathcal{F}_L \cap \mathcal{F}_M) = J(\mathcal{F}_L) \cap J(\mathcal{F}_M)
\]
\[
(2.7) \quad J(\mathcal{F}_L \cup \mathcal{F}_M) = J(\mathcal{F}_L) \cup J(\mathcal{F}_M).
\]

**Proposition 2.1.** Every \( L \in \mathcal{F} \) can be decomposed according to the members of \( J(\mathcal{F}) \) as follows:

\[
(2.8) \quad L = \bigcup \{ K | K \in J(\mathcal{F}_L) \}.
\]
Proof. Let \( K, M \in J(\mathfrak{X}) \) with \( K \neq M \), so that \( KM \subseteq K \) or \( KM \subseteq M \). Suppose that \( KM \subseteq M \). Then \( KM \subseteq \langle M \rangle \) and it follows that \([K] \cap [M] = K \cap \langle K \rangle \cap M \cap \langle M \rangle = \emptyset \), hence \( ([K]) | K \in J(\mathfrak{X}) \) is a disjoint family. The inclusion \( \mathfrak{X} \) in (2.8) is trivial. To establish \( \supseteq \) consider \( \iota \in L \). Define \( K_\iota := \cap \{ L' \in \mathfrak{X} | \iota \in L' \} \), the smallest set in \( \mathfrak{X} \) containing \( \iota \). Then \( K_\iota \in J(\mathfrak{X}) \), as seen from the following indirect argument. Suppose that \( K_\iota \notin J(\mathfrak{X}) \) and thus that \( K_\iota = L_1 \cup L_2 \) where \( L_1, L_2 \in \mathfrak{X} \), \( L_1 \subseteq K_\iota \), and \( L_2 \subseteq K_\iota \). Then \( \iota \in K_1 \) or \( \iota \in K_2 \) contradicting the minimality of \( K_\iota \). Finally, if \( \iota \in \langle K_\iota \rangle \) the minimality of \( K_\iota \) again would be contradicted, hence \( \iota \in [K_\iota] \). Since \( K_\iota \in J(\mathfrak{X}) \) this establishes the inclusion \( \supseteq \) in (2.8). \( \square \)

In particular, set \( L = I \) in (2.8) to obtain

\[
(2.9) \quad I = \hat{U}([K] | K \in J(\mathfrak{X})).
\]

For example, suppose that \( I = \{1, 2, 3\} \) and \( \mathfrak{X} \) is given by (1.3). Then \( J(\mathfrak{X}) \) is given by (1.7) and we find that \([\{1\}] = \{1\}, [\{1, 2\}] = \{2\}, \) and \([\{1, 3\}] = \{3\}, \) so (2.9) is evident.

2.2. The \( \mathfrak{X} \)-parameters of \( \Sigma \).

For any finite index sets \( T \) and \( U \) let \( \mathfrak{M}(T \times U) \) denote the vector space of all \( T \times U \) matrices, \( \mathfrak{P}(T) \) the cone of all positive definite \( T \times T \) matrices, \( \mathfrak{M}(T) \equiv \mathfrak{M}(T \times T) \) the algebra of all \( T \times T \) matrices, and \( \mathfrak{G}L(T) \) the group of all nonsingular \( T \times T \) matrices. For every \( \Sigma \in \mathfrak{P}(I) \) and every subset \( T \subseteq I \), let \( \Sigma_T \in \mathfrak{P}(T) \) denote the \( T \times T \) submatrix of \( \Sigma \) and let \( \Sigma_T^{-1} \) denote \( (\Sigma_T)^{-1} \). For \( K \in \mathfrak{X} \) partition \( \Sigma_K \) according to (2.3) as follows:
(2.10) \[
\Sigma_K = \begin{pmatrix} \Sigma_{<\mathcal{K}>} & \Sigma_{<\mathcal{K}>} \\ \Sigma_{[\mathcal{K}>} & \Sigma_{[\mathcal{K}>} \end{pmatrix},
\]

so \( \Sigma_{<\mathcal{K}>} \in \mathbb{P}(<\mathcal{K}>), \Sigma_{[\mathcal{K}>} \in \mathbb{P}([\mathcal{K}>]), \Sigma_{<\mathcal{K}>} \in \mathbb{M}([\mathcal{K}> \times <\mathcal{K}>), and \Sigma_{<\mathcal{K}>} = (\Sigma_{[\mathcal{K}>})^t.\)

Furthermore, define

(2.11) \[
\Sigma_{[\mathcal{K}>} \cdot \equiv \Sigma_{[\mathcal{K}>} \cdot <\mathcal{K}> = \Sigma_{[\mathcal{K}>} - \Sigma_{[\mathcal{K}>} \Sigma_{<\mathcal{K}>} \Sigma_{<\mathcal{K}>}^t \in \mathbb{P}(<\mathcal{K}>)
\]

and let \( \Sigma_{[\mathcal{K}>} \cdot \) denote \((\Sigma_{[\mathcal{K}>})^{-1}. Then for every \( x \in \mathbb{R}^I,\)

(2.12) \[
\text{tr}(\Sigma_{[\mathcal{K}>}^{-1}xx^t) = \\
\text{tr}(\Sigma_{[\mathcal{K}>}^{-1} \cdot <\mathcal{K}> \Sigma_{<\mathcal{K}>} <\mathcal{K}> \Sigma_{<\mathcal{K}>} x_{<\mathcal{K}>} x_{<\mathcal{K}>}^t) + \text{tr}(\Sigma_{<\mathcal{K}>}^{-1} <\mathcal{K}> <\mathcal{K}> x_{<\mathcal{K}>} x_{<\mathcal{K}>}^t).
\]

**Definition 2.2.** For \( \Sigma \in \mathbb{P}(I), \) the family of matrices

(2.13) \[
((\Sigma_{[\mathcal{K}>}^{-1} \cdot <\mathcal{K}>, \Sigma_{[\mathcal{K}>} \cdot )|K \in \mathcal{J}(\mathbb{X}))
\]

is called the family of \( \mathbb{X} \)-parameters of \( \Sigma.\)

\(\square\)

2.3. Characterization of conditional independence in terms of \( \Sigma^{-1}.\)

Theorem 2.1 presents an algebraic characterization of the set \( \mathbb{P}(\mathbb{X}) \) of covariance matrices \( \Sigma \) defined in terms of pairwise conditional independence (2.1). The following description of pairwise CI is useful.

**Lemma 2.1.** Let \( x \sim N(\Sigma), \Sigma \in \mathbb{P}(I). Then for any \( L, M \subseteq I, x_L \perp \!
\perp x_M \mid x_{L \cup M} \)

if and only if \( \forall x \in \mathbb{R}^I: \)

(2.14) \[
\text{tr}(\Sigma_{L \cup M}^{-1} x_{L \cup M} x_{L \cup M}^t) = \text{tr}(\Sigma_{L}^{-1} x_L x_L^t) + \text{tr}(\Sigma_{M}^{-1} x_M x_M^t) - \text{tr}(\Sigma_{L \cap M}^{-1} x_{L \cap M} x_{L \cap M}^t).\]
Proof. The difference

$$\text{tr}(\Sigma^{-1}_L x_L x_L^t) - \text{tr}(\Sigma^{-1}_M x_M x_M^t)$$

appears in the exponential term of the conditional density of

$$x_{(\cup M)} \setminus (\cup N)$$
given $$x_{\cup M}$$. Therefore $$x_L \perp x_M | x_{\cup M}$$ if and only if this difference is the sum of the differences appearing in the exponential terms of the conditional densities of $$x_{\cup (\cup M)}$$ given $$x_{\cup M}$$ and $$x_M \setminus (\cup M)$$ given $$x_{\cup M}$$. This sum is

$$(\text{tr}(\Sigma^{-1}_L x_L x_L^t) - \text{tr}(\Sigma^{-1}_L x_L x_M x_M^t)) + (\text{tr}(\Sigma^{-1}_M x_M x_M^t) - \text{tr}(\Sigma^{-1}_M x_M x_M^t)),$$

and the lemma follows. \(\square\)

Theorem 2.1. (Characterization of P(I).) For $$\Sigma \in P(I)$$ the following conditions are equivalent:

(i) $$\Sigma \in P(I)$$;

(ii) $$\forall x \in \mathbb{R}^I$$:

$$\text{tr}(\Sigma^{-1} xx^t) = \sum (\text{tr}(\Sigma^{-1}_K x_K x_K^t)(x_K - \Sigma^{-1}_K x_K <K>)^t) | K \in J(\mathfrak{L})$$;

(iii) $$\forall x \in \mathbb{R}^I, \forall L \in \mathfrak{L}$$:

$$\text{tr}(\Sigma^{-1}_L x_L x_L^t) = \sum (\text{tr}(\Sigma^{-1}_K x_K x_K^t)(x_K - \Sigma^{-1}_K x_K <K>)^t) | K \in J(\mathfrak{L}_L)$$.
Proof. Trivially (iii) => (ii). On the other hand, (iii) follows from (ii) if we replace \( I \) and \( \mathcal{A} \) by \( L \) and \( \mathcal{A}_L \), respectively, in (ii).

To show (i) => (ii), use induction on \( |J(\mathcal{A})| = q \). If \( q = 1 \) then by (2.4), \( \mathcal{A} = \{\emptyset, I\} \) and (ii) is trivial. Next, assume that (ii) is true whenever \( q \leq k-1 \) and suppose that \( q = k \). If \( I \in J(\mathcal{A}) \) then \( J(\mathcal{A}) = J(\mathcal{A}_{\langle I \rangle}) \hat{U} \{I\} \), hence \( |J(\mathcal{A}_{\langle I \rangle})| = k-1 \) and (iii) is true with \( L \) replaced by \( \langle I \rangle \), so (ii) follows from (2.12) with \( K \) replaced by \( I \). If, on the other hand, \( I \notin J(\mathcal{A}) \), then \( I = LM \) where \( L \subset I \) and \( M \subset I \). It follows from (2.4) that \( |J(\mathcal{A}_L)| < k \) and \( |J(\mathcal{A}_M)| < k \), so by the induction assumption, (iii) is valid with \( L \) replaced by \( L, M, \) and \( LM \). Then (ii) follows from (2.6), (2.7), and Lemma 2.1.

To show (iii) => (i), consider any pair \( L, M \in \mathcal{A} \). Apply condition (iii) four times, with \( L \) replaced by \( LM \), \( L \), \( M \), and \( LM \), and then apply (2.6) and (2.7) to obtain (2.14). By Lemma 2.1, therefore, (i) is satisfied.

2.4. The \( \mathcal{A} \)-preserving matrices: generalized block-triangular matrices with lattice structure.

We now introduce a group \( \text{GL}_\mathcal{A}(I) \) of nonsingular matrices \( A \) that will be seen in Section 2.6 to act transitively on \( P_\mathcal{A}(I) \). In the present section \( \text{GL}_\mathcal{A}(I) \) is shown to be a group of block-triangular matrices with lattice structure determined by \( \mathcal{A} \).

For any \( A \in \text{M}(I) \) and any two subsets \( L, M \in J(\mathcal{A}) \) let \( A_{[LM]} \) denote the \([L] \times [M] \) submatrix of \( A \).

**Proposition 2.2.** Let \( A \in \text{M}(I) \). The following three conditions on \( A \) are equivalent:
(i) \( \forall x \in \mathbb{R}^I, \forall L \in \mathcal{I}: x_L = 0 \Rightarrow (Ax)_L = 0; \)

(ii) \( \forall x \in \mathbb{R}^I, \forall L \in \mathcal{I}: (Ax)_L = A_L x_L; \)

(iii) \( \forall L, M \in J(\mathcal{I}): M \not\subseteq L \Rightarrow A_{[LM]} = 0. \)

Proof: (ii) \( \Rightarrow \) (i) is trivial.

(iii) \( \Rightarrow \) (ii): By the usual formula for matrix multiplication by blocks,

\[
(Ax)_L = \sum_{M \in J(\mathcal{I})} (A_{[KM]} x_{[M]} | M \in J(\mathcal{I})) | K \in J(x_L))
= \sum_{M \in J(\mathcal{I})} (A_{[KM]} x_{[M]} | M \in J(\mathcal{I})) | K \in J(x_L))
= A_L x_L.
\]

The first equality uses (2.8) and (2.9), the second uses condition (i), while the third uses (2.8) twice.

(i) \( \Rightarrow \) (iii): Suppose \( L, M \in J(\mathcal{I}) \) with \( M \not\subseteq L \). Let \( \epsilon \) denote any column vector in \( \mathbb{R}^I \) satisfying \( \epsilon_{[K]} = 0 \) for \( K \in J(\mathcal{I}), K \neq M \). Then

\[
A_{[LM]} \epsilon_{[M]} = \sum_{K \in J(\mathcal{I})} (A_{[LK]} \epsilon_{[K]} | K \in J(\mathcal{I})) = (A \epsilon)_{[L]}. \]

But \( (A \epsilon)_L = 0 \) by (i), hence \( (A \epsilon)_{[L]} = 0 \). Since \( \epsilon_{[M]} \) is arbitrary this implies \( A_{[LM]} = 0 \) as required. \( \square \)

Let \( M_\mathcal{A}(I) \) denote the set of all \( A \in M(I) \) that satisfy the equivalent conditions (i), (ii), (iii) in Proposition 2.2 and let \( GL_\mathcal{A}(I) \) denote the set of all nonsingular matrices in \( M_\mathcal{A}(I) \). It follows from (i) that \( M_\mathcal{A}(I) \) is a matrix algebra and hence \( GL_\mathcal{A}(I) \) is a matrix group. It also follows by (i) that \( M_\mathcal{A}(I) \) is the set of all matrices that, for each \( L \in \mathcal{I}, \)
preserve the kernel of the projection $\mathbb{R}^I \rightarrow \mathbb{R}^L$ given by $x \rightarrow x_L$. Note that when $\mathcal{X} = \{\emptyset, I\}$, $M_\mathcal{X}(I) = M(I)$ and $GL_\mathcal{X}(I) = GL(I)$.

**Definition 2.3.** The algebra $M_\mathcal{X}(I)$ is called the algebra of $\mathcal{X}$-preserving matrices and $GL_\mathcal{X}(I)$ the group of $\mathcal{X}$-preserving matrices. $\square$

**Remark 2.1.** When $\mathcal{X}$ is a chain then $J(\mathcal{X}) = \mathcal{X}\setminus\{\emptyset\}$ is also a chain, so it follows from Proposition 2.2 (iii) that $M_\mathcal{X}(I)$ is an algebra of block-triangular matrices in the usual sense. For a general $\mathcal{X}$ let $q := |J(\mathcal{X})|$ and let $K_1, K_2, \ldots, K_q$ be a never-decreasing listing of the members of the poset $J(\mathcal{X})$, i.e., $i < j \Rightarrow K_j \not\subseteq K_i$. If every $A \in M(I)$ is partitioned according to the ordered decomposition

$$I = [K_1] \cup [K_2] \cup \cdots \cup [K_q],$$

then it is seen from Proposition 2.2 (iii) that $M_\mathcal{X}(I)$ can be represented as a subalgebra of the algebra of lower block-triangular matrices. That is, $A \in M_\mathcal{X}(I)$ is lower block-triangular with additional blocks of zeroes below the main diagonal - see (1.13) and also Section 2.8 for further examples. $\square$

**Remark 2.2.** For $K \in \mathcal{X}$ and $A \in M(I)$ let $A_K$ denote the $K\times K$ submatrix of $A$ and partition $A$ according to (2.3) and (2.10) as follows:

$$A_K = \begin{bmatrix} A_{\prec K} & A_{\ll K} \\ A_K & A_{\succ K} \end{bmatrix}.$$
note that $A_{[KK]} = A_{[K]}$ when $K \in J(\mathcal{X})$. By Proposition 2.2(ii), if $A \in M_{\mathcal{X}}(I)$ then for every $K \in J(\mathcal{X})$ and $x \in \mathbb{R}^I$,

\begin{align*}
(2.17) & \quad A_{[K]} = 0 \\
(2.18) & \quad (Ax)[K] = A_{[K]}x[K] + A_{[K]}x(K)^{<K>}.
\end{align*}

Furthermore, the linear mapping

\begin{align*}
(2.19) & \quad M_{\mathcal{X}}(I) \rightarrow X(M([K]^{<K>}) \times M([K]) | K \in J(\mathcal{X})) \\
& \quad A \rightarrow ((A_{[K]}^{[K]}, A_{[K]}) | K \in J(\mathcal{X}))
\end{align*}

is bijective. This holds because, by Proposition 2.2(iii), $A \in M_{\mathcal{X}}(I)$ if and only if the $[K] \times (I \setminus K)$-submatrix of $A$ is 0 for every $K \in J(\mathcal{X})$. Under the correspondence (2.19) the subset $GL_{\mathcal{X}}(I)$ corresponds to the subset

\begin{align*}
(2.20) & \quad X(M([K]^{<K>}) \times GL([K]) | K \in J(\mathcal{X})). \quad \Box
\end{align*}

**Lemma 2.2.** For $A \in GL_{\mathcal{X}}(I)$, $L \in \mathcal{X}$, and $K \in J(\mathcal{X})$,

\begin{align*}
(2.21) & \quad (A^{-1})_L = (A_L)^{-1} =: A_L^{-1} \\
(2.22) & \quad (A^{-1})[K] = (A[K])^{-1} =: A^{-1}_{[K]} \\
(2.23) & \quad (A^{-1})_{[K]}A^{<K>} = - A^{-1}_{[K]}A_{[K]}^{<K>}.
\end{align*}

**Proof.** From Proposition 2.2(ii), $(AC)_L = A_L C_L$ for every $A, C \in M_{\mathcal{X}}(I), L \in \mathcal{X}$, which implies (2.21). Then (2.22) and (2.23) follow from (2.17). \( \Box \)
Lemma 2.3. The mapping

\[ P(I) \to X(M([K] \times [K]) \times P([K]) | K \in J(\mathfrak{A})) \]

\[ \Sigma \to ((\Sigma_{[K]}^{-1}, \Sigma_{[K]}^{t}) | K \in J(\mathfrak{A})) \]

from \( \Sigma \) to its \( \mathfrak{A} \)-parameters commutes with the actions of \( GL_{\mathfrak{A}}(I) \) on \( P(I) \) and on \( X(M([K] \times [K]) \times P([K]) | K \in J(\mathfrak{A})) \) given by

\[ GL_{\mathfrak{A}}(I) \times P(I) \to P(I) \]

\[ (A, \Sigma) \to A\Sigma A^t \]

and

\[ GL_{\mathfrak{A}}(I) \times (X(M([K] \times [K]) \times P([K]) | K \in J(\mathfrak{A})) \]

\[ \to X(M([K] \times [K]) \times P([K]) | K \in J(\mathfrak{A})) \]

\[ (A, (R_{[K]}, A_{[K]}) | K \in J(\mathfrak{A})) \]

\[ \to ((A_{[K]} A_{[K]}^{-1} + A_{[K]} A_{[K]}^{-1}, A_{[K]} A_{[K]} A_{[K]}^t) | K \in J(\mathfrak{A})). \]

respectively.

Proof. It is straightforward to verify that (2.26) is a group action. We must show that for every \( A \in GL_{\mathfrak{A}}(I) \), \( \Sigma \in P(I) \), and \( K \in J(\mathfrak{A}) \).

\[ (A\Sigma A^t)_{[K]} (A\Sigma A^t)_{[K]}^{-1} = A_{[K]} \Sigma_{[K]} A_{[K]}^{-1} + A_{[K]} A_{[K]}^{-1} \]

and

\[ (A\Sigma A^t)_{[K]} = A_{[K]} \Sigma_{[K]} A_{[K]}^t. \]

It follows from Proposition 2.2(ii) that \( (A\Sigma A^t)_{K} = A_{K} \Sigma_{K} A_{K}^t \). Let \( A_{K} \) and \( \Sigma_{K} \) be partitioned as in (2.16) and (2.10), respectively. Since \( A_{[K]} = 0 \), (2.27) and (2.28) follow by direct calculation. \( \square \)
Proposition 2.3. If $\Sigma \in P_{\mathcal{A}}(I)$ and $\Lambda \in \mathcal{G}_{\mathcal{A}}(I)$, then $\Lambda \Sigma \Lambda^t \in P_{\mathcal{A}}(I)$.

Proof. We shall show that condition (ii) of Theorem 2.1 is valid with $\Sigma$ replaced by $\Lambda \Sigma \Lambda^t$. Since $\Sigma \in P_{\mathcal{A}}(I)$, (ii) holds for $\Sigma$. Now replace $x$ by $\Lambda^{-1}x$ in (ii) and let $B = \Lambda^{-1}$. The left-hand side of (ii) becomes

$$\text{tr}((\Lambda \Sigma \Lambda^t)^{-1}xx^t)$$

while the summands on the right-hand side become

$$\text{tr}((\Sigma[K])^{-1}[K](Bx)[K] - \Sigma[K]\Sigma^{-1}<K><K>(Bx)<K>(\cdots)^t)$$

= \text{tr}((\Sigma[K])^{-1}[K](Bx)[K] + B[K]<K> - \Sigma[K]\Sigma^{-1}<K><K>B[K]<K>)(\cdots)^t)

= \text{tr}((\Lambda^{-1}[K]\Sigma[K])^{-1}[K](Bx)[K] - (\Lambda^{-1}[K]\Sigma[K])^{-1}[K]<K>B[K]<K>)(\cdots)^t)

= \text{tr}((\Lambda^{-1}[K]\Sigma[K])^{-1}[K](Bx)[K] - (\Lambda^{-1}[K]\Sigma[K])^{-1}[K]<K>x[K])(\cdots)^t).

The first equality uses (2.18) and Proposition 2.2(ii), the third uses (2.22) and (2.23), and the fourth uses (2.27) and (2.28). Therefore condition (ii) of Theorem 2.1 holds for $\Lambda \Sigma \Lambda^t$. □

2.5. The $\mathcal{A}$-parametrization of $P_{\mathcal{A}}(I)$.

Theorem 2.2 below establishes the one-to-one correspondence between $\Sigma$ and its $\mathcal{A}$-parameters. Together with Theorem 2.1(ii) and Lemma 2.5, this decomposition of the parameter space $P_{\mathcal{A}}(I)$ yields the fundamental factorization of the likelihood function for the CI model $\mathcal{N}(\mathcal{A})$ (cf. Theorem 3.1).

Lemma 2.4. For any family

$$((R^K, A^K) | K \in J(\mathcal{A})) \in \mathcal{X}(\mathcal{A}[K]<K> \times \mathcal{P}(K) | K \in J(\mathcal{A}))$$
there exists a matrix $A \in \text{GL}_\mathcal{A}(I)$ such that for every $K \in J(\mathcal{A})$.

(2.29) \[ A_{\langle K \rangle} = R_{[K]} A_{\langle K \rangle} \]

(2.30) \[ A_{[K]} A_{[K]}^t = A_{[K]} \]

Proof. First choose matrices $A_{[K]} \in \text{GL}([K]), K \in J(\mathcal{A})$, that satisfy (2.30). As in Remark 2.1 let $K_1, \ldots, K_q$ be a never-decreasing listing of the elements in $J(\mathcal{A})$. For notational convenience abbreviate $K_k$ by $k$, $\langle K_k \rangle$ by $\langle k \rangle$, $[K_k]$ by $[k]$, and $[K_k]$ by $[k]$ whenever they appear as subscripts.

If $K_1 \subset K_2$ then $\langle K_2 \rangle = [K_1]$, so $A_{\langle 2 \rangle} = A_{[1]}$ and $A_{\langle 2 \rangle}$ is uniquely determined by (2.29); if $K_1 \notin K_2$ then $\langle K_2 \rangle = \emptyset$ so (2.29) is vacuous. Now suppose that we have determined $A_{\langle 2 \rangle}, \ldots, A_{\langle k-1 \rangle}$ satisfying (2.29). These $k-2$ matrices (some of which may be vacuous), together with $A_{[1]}, \ldots, A_{[k-1]}$, completely determine $A_{\langle k \rangle}$. This follows from the decomposition (cf. (2.8))

(2.31) \[ \langle k \rangle \equiv \langle K_k \rangle = \hat{\cup}([K_1]|K_1 \subset \langle K_k \rangle) \]

and the fact that $K_1 \subset \langle K_k \rangle \Rightarrow i < k$ for a never-decreasing listing. Now $A_{\langle k \rangle}$ is uniquely determined by (2.29) and, after induction on $k$, the matrix $A$ is completely determined. By the surjectivity of the mapping (2.19), $A \in \text{GL}_\mathcal{A}(I)$. \hfill \Box

Theorem 2.2. (The $\mathcal{A}$-parametrization of $P_{\mathcal{A}}(I)$.) The following mapping is bijective:
Proof. By Theorem 2.2(ii), (2.32) is injective. To show that (2.32) is surjective, consider

$$\left\{(R^{[K]}, A^{[K]} \mid K \in J(\mathcal{A})) \in X(M([K] \times [K] \times P([K]) \mid K \in J(\mathcal{A})) \right\}.$$

By Lemma 2.4 there exists a matrix $A \in \text{GL}_{\mathcal{A}}(I)$ satisfying (2.29) and (2.30). Define $\Sigma := AA^t$; then $\Sigma \in P_{\mathcal{A}}(I)$ by Proposition 2.3 (with $\Sigma = 1_I$).

The $\mathcal{X}$-parameters of $\Sigma$ are given by $\Sigma^{[K]}_{[K]} = A^{[K]}A^{-1}_{[K]} = R^{[K]}$ and $\Sigma^{[K]}_{[K]} = A^{[K]}A^{t}_{[K]} = A^{[K]}$, $K \in J(\mathcal{A})$ (set $\Sigma = 1_I$ in (2.27) and (2.28)). $\square$

2.6. Transitive action of the group of $\mathcal{X}$-preserving matrices.

Theorem 2.3. The action

$$(2.33) \quad \text{GL}_{\mathcal{A}}(I) \times P_{\mathcal{A}}(I) \to P_{\mathcal{A}}(I)$$

$$(A, \Sigma) \to A\Sigma A^t$$

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

Proof: That (2.33) is well-defined follows from Proposition 2.3. By Lemma 2.3, the bijective mapping (2.32) commutes with the actions (2.33) and (2.26). By Lemma 2.4, however, the action (2.26) is transitive, so it follows that (2.33) also is transitive. That (2.33) is continuous is trivial. Since $P_{\mathcal{A}}(I)$ and $\text{GL}_{\mathcal{A}}(I)$ are closed subsets of $P(I)$ and $\text{GL}(I)$, respectively, and the classical action of $\text{GL}(I)$ on $P(I)$ is proper, it follows that the action (2.33) also is proper. $\square$
Remark 2.3. Set $P^\mathcal{A}(I)^{-1} := \{ A \in P(I) | A^{-1} \in P^\mathcal{A}(I) \}$. By Theorem 2.3, the action

\[(2.34) \quad \text{GL}^\mathcal{A}(I) \times P^\mathcal{A}(I)^{-1} \to P^\mathcal{A}(I)^{-1} \]
\[(A, A^{-1}) \mapsto (A^{-1})^t \Delta A^{-1} \]

induced on $P^\mathcal{A}(I)^{-1}$ by (2.33) is also well-defined, transitive, continuous, and proper. □

Remark 2.4. Since both $P^\mathcal{A}(I)$ and $P^\mathcal{A}(I)^{-1}$ contain the $I \times I$ identity matrix $I$, it follows from the transitivity of the actions (2.33) and (2.34) that

\[(2.35) \quad P^\mathcal{A}(I) = \{ AA^t \in P(I) | A \in \text{GL}^\mathcal{A}(I) \} \]
\[(2.36) \quad P^\mathcal{A}(I)^{-1} = \{ A^t A \in P(I) | A \in \text{GL}^\mathcal{A}(I) \}. \]

If $\mathcal{A} = \emptyset, I$ then $P^\mathcal{A}(I) = P^\mathcal{A}(I)^{-1} = P(I)$, so both actions (2.33) and (2.34) reduce to the well-known transitive actions of $\text{GL}(I)$ on $P(I)$. If $\mathcal{A}$ is a chain as in Examples 2.1 and 2.2 in Section 2.8 then again $P^\mathcal{A}(I) = P^\mathcal{A}(I)^{-1} = P(I)$, but now $\text{GL}^\mathcal{A}(I)$ is a group of nonsingular lower block-triangular matrices in the usual sense and the actions (2.33) and (2.34) are the well-known transitive actions of $\text{GL}^\mathcal{A}(I)$ on $P(I)$.

The following lemma generalizes the Schur decomposition formula for $\det(\Sigma)$. 
Lemma 2.5. For $\Sigma \in P_{\Delta}(I)$,

\begin{equation}
\det(\Sigma) = \Pi(\det(\Sigma_{[K],}) | K \in \mathcal{J}(\mathcal{A})).
\end{equation}

**Proof:** By Theorem 2.3 there exists $A \in GL_{\Delta}(I)$ such that $\Sigma = AA^t$. Thus

\[
\det(\Sigma) = \det(AA^t) = \Pi(\det(A[IK]A^t[IK]) | K \in \mathcal{J}(\mathcal{A})) = \Pi(\det(\Sigma_{[K],}) | K \in \mathcal{J}(\mathcal{A})).
\]

The second equality holds since $A$ can be represented as a lower block-triangular matrix (cf. Remark 2.1), while the third equality follows from (2.29).

\[\Box\]

2.7. Reconstruction of $\Sigma$ from its $\mathcal{A}$-parameters.

By Theorem 2.2, $\Sigma \in P_{\Delta}(I)$ is uniquely determined by its $\mathcal{A}$-parameters

\[
((R_{[K]}, \Lambda_{[K]}) | K \in \mathcal{J}(\mathcal{A})) \in \mathcal{X}(\mathcal{M}(\mathcal{K} \times \mathcal{K}) \times \mathcal{P}(\mathcal{K}) | K \in \mathcal{J}(\mathcal{A})),
\]

where

\begin{equation}
R_{[K]} = \Sigma_{[K]} \Sigma_{[K]}^{-1} \quad \text{and} \quad \Lambda_{[K]} = \Sigma_{[K]}.
\end{equation}

Because the MLE $\hat{\Sigma}$ is obtained by first estimating the $\mathcal{A}$-parameters of $\Sigma$, then using these estimates to obtain $\hat{\Delta}$ itself, it is important to find an explicit method for reconstructing $\Sigma \in P_{\Delta}(I)$ from its $\mathcal{A}$-parameters.

One such method is to apply the formula

\begin{equation}
\Sigma^{-1} = \sum_{K} (\Lambda_1(K) | K \in \mathcal{J}(\mathcal{A})).
\end{equation}
which is just a re-expression of Theorem 2.1(ii), where $\Lambda_1(K)$ is the $I \times I$ matrix whose $K \times K$ submatrix is

$$
(2.40) \begin{pmatrix}
R^t_{[K]A^{-1}[K]} R_{[K]} & -R^t_{[K]A^{-1}[K]} \\
-A^{-1}_{[K]} R_{[K]} & A_{[K]}
\end{pmatrix}
$$

and whose remaining entries are 0. In general, however, it is not a simple task to determine $\Sigma$ from (2.39) by matrix inversion. We now present a step-wise algorithm for reconstructing $\Sigma$ directly from its $\mathcal{X}$-parameters.

Let $K_1, \ldots, K_q$ be a never-decreasing listing of the members of the poset $J(\mathcal{X})$ (cf. Remark 2.1 and the proof of Lemma 2.4). partition $\Sigma$ according to (2.9), and list the $\mathcal{X}$-parameters in the corresponding order:

$$
(2.41) \quad (\Lambda_{[1]}, R_{[2]A_{[2]}}, \ldots, R_{[q]A_{[q]}}) \in P([K_1]) \times M([K_2] \times [K_2]) \times \cdots \times M([K_q] \times [K_q]) \times P([K_q]).
$$

(Recall that whenever they appear as subscripts, $K_k$, $\langle K_k \rangle$, $[K_k]$, and $[K_k]$ are abbreviated by $k$, $\langle k \rangle$, $[k]$, and $[k]$, respectively.) The reconstruction algorithm proceeds step-wise as follows. At step $k$ the relations in (2.38) are inverted to determine $\Sigma_{[k]}$ and $\Sigma_{[k]}$. from the corresponding $\mathcal{X}$-parameters $R_{[k]}$ and $A_{[K]}$ and from the matrix $\Sigma_{1 \cup \cdots \cup (k-1)}$ constructed in step $k-1$. The remaining entries in $\Sigma_{1 \cup \cdots \cup k}$ are determined by the CI conditions.

**Step 1:**

$$
\Sigma_{[1]} = A_{[1]}.
$$
Step 2: 
\[ \Sigma_{[2]} = R_{[2]} \Sigma_{<2>} \]
\[ \Sigma_{[2]} = \Lambda_{[2]} + R_{[2]} \Sigma_{<2>} \]

At this point the submatrix \( \Sigma_{1\cup2} \) is completely determined: if \( K_1 \subseteq K_2 \) then \( \Sigma_{1\cup2} = \Sigma_2 \), while if \( K_1 \not\subseteq K_2 \) then \( K_1 \cup K_2 = \emptyset \) so the \([K_1]_{x}[K_2]_{-}\)-submatrix of \( \Sigma \) is 0 by (2.2). (Recall that \( 1\cup2 \) abbreviates \( K_1 \cup K_2 \) when appearing as a subscript.) By (2.42), \( \langle K_3 \rangle \subseteq K_1 \cup K_2 \), so \( \Sigma_{<3>} \) is a submatrix of \( \Sigma_{1\cup2} \), hence the next step may be carried out.

Step 3a: 
\[ \Sigma_{[3]} = R_{[3]} \Sigma_{<3>} \]
\[ \Sigma_{[3]} = \Lambda_{[3]} + R_{[3]} \Sigma_{<3>} \]

It is important to note that after Steps 1, 2, and 3a, the three submatrices \( \Sigma_1, \Sigma_2, \Sigma_3 \) are now determined but the complete submatrix \( \Sigma_{1\cup2\cup3} \) may not yet be fully determined. The remaining \([K_3]_{x}((K_1 \cup K_2 \cup K_3) \setminus K_3)\)-submatrix of \( \Sigma_{1\cup2\cup3} \), which we denote by \( \Sigma_{<3>} \), is determined from \( \Sigma_{1\cup2} \) by means of the pairwise CI requirements imposed by \( \mathcal{A} \) (cf. (2.44)).

Step 3b: 
\[ \Sigma_{[3]} = R_{[3]} \Sigma_{<3>} \]
\[ (= \Sigma_{[3]} \Sigma^{-1}_{<3} \Sigma_{<3>} ) \]

where \( \Sigma_{<3>} \) is the \( \langle K_3 \rangle_{x}((K_1 \cup K_2 \cup K_3) \setminus K_3)\)-submatrix of \( \Sigma_{1\cup2\cup3} \). By (2.42) and (2.43), however, \( \Sigma_{<3>} \) is in fact a submatrix of \( \Sigma_{1\cup2} \), hence may be used to obtain \( \Sigma_{[3]} \) in this step.

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

$$K_1 U \cdots U K_k = \hat{U}([K_j]_{j=1, \cdots, k}),$$

$$K_k = \hat{U}([K_j]_{j=1, \cdots, k}, K_j \subseteq K_k).$$

From these relations and (2.3) it may be deduced that

$$(2.42) \quad \langle K_k \rangle = K_k \cap (K_1 U \cdots U K_{k-1}) \subseteq K_1 U \cdots U K_{k-1}$$

$$(2.43) \quad (K_1 U \cdots U K_k) \setminus K_k = (K_1 U \cdots U K_{k-1}) \langle K_k \rangle \subseteq K_1 U \cdots U K_{k-1}.$$

Thus, if we denote the $[K_k]_{x ((K_1 U \cdots U K_k) \setminus K_k)}$-submatrix of $\Sigma_{1U \cdots U_k}$ by $\Sigma_{\langle k \rangle}$ and the $\langle K_k \rangle \times ((K_1 U \cdots U K_k) \setminus K_k)$-submatrix by $\Sigma_{\langle k \rangle}$, it follows from (2.42) and (2.43) that both $\Sigma_{\langle k \rangle}$ and $\Sigma_{\langle k \rangle}$ are in fact submatrices of $\Sigma_{1U \cdots U (k-1)}$, so the next step may be carried out:

**Step k:**

$$\Sigma_{\langle k \rangle} = R_{\langle k \rangle} \Sigma_{\langle k \rangle},$$

$$\Sigma_{[k]} = A_{[k]} + R_{\langle k \rangle} \Sigma_{\langle k \rangle},$$

$$(2.42)\quad \Sigma_{[k]} = R_{\langle k \rangle} \Sigma_{\langle k \rangle}$$

$$= \Sigma_{\langle k \rangle}^{-1} \Sigma_{\langle k \rangle}.$$}

The relation in (2.44) is seen as follows. Since $K_k (=: L)$ and $K_1 U \cdots U K_{k-1} (=: M)$ are members of $\mathcal{M}$ it follows from the pairwise CI condition (2.2) and from (2.42) and (2.43) that the $[K_k]_{x ((K_1 U \cdots U K_k) \setminus K_k)}$-submatrix of $(\Sigma_{1U \cdots U_k})^{-1}$ is a zero matrix, which is equivalent to (2.44).

The submatrix $\Sigma_{1U \cdots U_k}$ of $\Sigma$ is fully determined after Step k; after q steps, $\Sigma_{1U \cdots U_k} \equiv \Sigma$ 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.]

2.8. Examples.

A series of nine Examples will illustrate the following basic aspects of a lattice CI model \( \mathcal{X}(\mathcal{I}) \): (a) the distributive lattice \( \mathcal{X} \subseteq \mathcal{I}(\mathcal{I}) \) and the poset \( J(\mathcal{X}) \) of join-irreducible elements; (b) the \( \mathcal{X} \)-parametrization (2.32) of \( \mathcal{P}_{\mathcal{X}}(\mathcal{I}) \) and the associated decomposition of \( \text{tr}(\Sigma^{-1} xx^t) \) given in Theorem 2.1(ii); (c) the choice of a never-decreasing listing of the members of \( J(\mathcal{X}) \) and the reconstruction of the covariance matrix \( \Sigma \in \mathcal{P}_{\mathcal{X}}(\mathcal{I}) \) from its ordered \( \mathcal{X} \)-parameters (cf. (2.38)) by means of the step-wise algorithm in Section 2.7, as well as the form of the precision matrix \( \Lambda = \Sigma^{-1} \in \mathcal{P}_{\mathcal{X}}(\mathcal{I})^{-1} \); (d) the form of the \( \mathcal{X} \)-preserving matrices, i.e., the group \( \text{GL}_{\mathcal{X}}(\mathcal{I}) \) of matrices, partitioned according the ordered decomposition (2.15), that acts transitively on \( \mathcal{P}_{\mathcal{X}}(\mathcal{I}) \) (cf. Remarks 2.1 and 2.2). The reader should verify directly that (2.35) and (2.36) hold for \( \mathcal{P}_{\mathcal{X}}(\mathcal{I}) \) and \( \text{GL}_{\mathcal{X}}(\mathcal{I}) \) in these nine Examples.

In each Example the lattice diagram of \( \mathcal{X} \) appears in an accompanying Figure, in which the members of \( J(\mathcal{X}) \) are indicated by open circles and the remaining members of \( \mathcal{X} \) by solid dots. In each Figure the minimal element \( \emptyset \) appears at the left while the maximal element \( I \) appears at the right.

These Examples will be continued in Section 3.2, where the MLE \( \hat{\Lambda} \) is determined for each of these CI models, and in Section 4.3 to provide examples of the problem of testing one CI model against another. Additional examples appear in [AP] (1991).
Example 2.1. First consider the simple case where $\mathcal{X} = \{\emptyset, L, I\}$ (see Figure 2.1).

![Figure 2.1](image)

Since $\mathcal{X}$ is a chain, $P_{\mathcal{X}}(I) = P(I)$. Note that $J(\mathcal{X}) = \{L, I\}$ and $\langle L \rangle = \{0\}$, $\langle I \rangle = [L] = L$. Thus the $\mathcal{X}$-parametrization of $P_{\mathcal{X}}(I)$ becomes

\[
P(I) \leftrightarrow P(L) \times M([I] \times L) \times P([I])
\]

and

\[
\Sigma \leftrightarrow (\Sigma_L, \Sigma_{[I]}^{\Sigma_L^{-1}}, \Sigma_{[I]}).
\]

The algorithm for reconstructing $\Sigma$ from its ordered $\mathcal{X}$-parameters $A_{[L]}$, $R_{[I]}$, $A_{[I]}$ takes the following form:

**Step 1:**

\[
\Sigma_L = A_{[L]}
\]

**Step 2:**

\[
\Sigma_{[I]} = R_{[I]} \Sigma_L \Sigma_{[I]} = A_{[I]} + R_{[I]} \Sigma_{[I]}
\]

The group $GL_{\mathcal{X}}(I)$ is a lower block-triangular matrix group in the ordinary sense: $GL_{\mathcal{X}}(I)$ consists of all nonsingular $I \times I$ matrices of the form:

\[
A = \begin{bmatrix}
A_L & 0 \\
A_{[L]} & A_{[I]}
\end{bmatrix}
\]
Example 2.2. If $\mathcal{A} \equiv \{\emptyset \equiv K_0, K_1, \ldots, K_{q-1}, K \equiv I\}$ is an ascending chain, i.e., $\emptyset \subset K_1 \subset \cdots \subset K_{q-1} \subset I$, then a well-known generalization of the preceding example is obtained (see Figure 2.2).

\[
\emptyset \longrightarrow \cdots \longrightarrow I
\]

\[
\begin{array}{cc}
K_1 & K_{q-1}
\end{array}
\]

Figure 2.2.

Again $P_{\mathcal{A}}(I) = P(I)$, but the $\mathcal{A}$-parametrization is changed. Note that $J(\mathcal{A}) = \{K_1, \ldots, K_q\}$ and $\langle K_1 \rangle = \emptyset$, $\langle K_k \rangle = K_{k-1}$, $k = 2, \ldots, q$. Then the $\mathcal{A}$-parametrization of $P_{\mathcal{A}}(I)$ becomes

\[
P(I) \leftrightarrow P(K_1) \times M([K_2] \times K_1) \times P([K_2]) \times \cdots \times M([K_q] \times K_{q-1}) \times P([K_q])
\]

$\Sigma \leftrightarrow (\Sigma_1, \Sigma_2, \Sigma_1^{-1}, \Sigma_2, \ldots, \Sigma_q, \Sigma_1^{-1}, \Sigma_q, \ldots, \Sigma_{q-1})$.

and

\[
tr(\Sigma^{-1}xx^t) = tr(\Sigma_1^{-1}x_1x_1^t) + tr(\Sigma_2^{-1}(x_2 - \Sigma_2^{-1}x_1)(\cdots)^t) + \cdots + tr(\Sigma_q^{-1}(x_q - \Sigma_q^{-1}x_{q-1})(\cdots)^t).
\]

where $K_1, K_2, \ldots, K_q$ are abbreviated as $1, 2, \ldots, q$ whenever they occur as subscripts. Then $\Sigma$ is reconstructed from its ordered $\mathcal{A}$-parameters $\Lambda_1,$ $\Lambda_2$, $\ldots$, $\Lambda_q$, as follows:

**Step 1:**

$\Sigma_1 = \Lambda_1$

**Step 2:**

$\Sigma_2 = \Lambda_2 + R_{[2]}^2 \Sigma_1$

$\vdots$

**Step q:**

$\Sigma_q = \Lambda_q + R_{[q]}^q \Sigma_{q-1}$

$\Sigma_q = \Lambda_q + R_{[q]}^q \Sigma_{q-1}$.
The group $GL_{\mathcal{A}}(I)$ is again a group of lower block-triangular matrices in the usual sense. For example, when $q = 4$, $GL_{\mathcal{A}}(I)$ consists of all nonsingular $I \times I$ matrices of the form

$$
A = \begin{bmatrix}
A_1 & 0 & 0 & 0 \\
A_2 & 0 & 0 & 0 \\
A_3 & 0 & 0 & 0 \\
A_4 & 0 & 0 & 0 \\
\end{bmatrix}
$$

(2.50)

Example 2.3. Consider the lattice $\mathcal{A} = \{\emptyset \equiv L \cap M, L, M, L \cap M \equiv I\}$ (see Figure 2.3).

Here the CI requirement determined by $\mathcal{A}$ is nontrivial, so $P_{\mathcal{A}}(I) \subset P(I)$.

Now $J(\mathcal{A}) = \{L, M\}$ and $\langle L \rangle = \langle M \rangle = \emptyset$. The $\mathcal{A}$-parametrization takes the form

$$
P_{\mathcal{A}}(I) \leftrightarrow P(L) \times P(M)$$

(2.51)

$$\Sigma \leftrightarrow (\Sigma_L, \Sigma_M),$$

and

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

(2.52)

Since $L, M$ is a never-decreasing listing of $J(\mathcal{A})$, $\Sigma$ may be reconstructed from its ordered nontrivial $\mathcal{A}$-parameters $A_{[L]}, A_{[M]}$ as follows:
Step 1: \[ \Sigma_L = A_L [L] \]
Step 2: \[ \Sigma_M = A_M [M] \]
\[ \Sigma_{[M]} = 0. \]

Thus \( \mathcal{P}_\mathcal{A}(I) \) consists of all block-diagonal matrices \( \Sigma \) of the form

\begin{equation}
\Sigma = \begin{bmatrix} \Sigma_L & 0 \\ 0 & \Sigma_M \end{bmatrix}
\end{equation}

where \( \Sigma \) is partitioned according to the ordered decomposition

\begin{equation}
I = L \cup M.
\end{equation}

In this Example, as in Examples 2.1 and 2.2, \( \mathcal{P}_\mathcal{A}(I) = \mathcal{P}_\mathcal{A}(I)^{-1} \) and both are linear, i.e., closed under (nonnegative) linear combinations. The group \( \mathcal{G}_\mathcal{A}(I) \) consists of all nonsingular \( I \times I \) matrices of the form

\begin{equation}
A = \begin{bmatrix} A_L & 0 \\ 0 & A_M \end{bmatrix}
\end{equation}

Example 2.4. If \( \mathcal{A} = \{ \emptyset \equiv L \cup M, L, M, L \cup M, I \} \) (see Figure 2.4)

![Figure 2.4](image)

then again \( \mathcal{P}_\mathcal{A}(I) \subset \mathcal{P}(I) \). Here \( \mathcal{J}(\mathcal{A}) = \{L, M, I\} \) and \( \langle L \rangle = \langle M \rangle = \emptyset, \langle I \rangle = L \cup M \). The \( \mathcal{A} \)-parametrization of \( \mathcal{P}_\mathcal{A}(I) \) assumes the form
(2.56) \[ \mathcal{P}_I(I) \leftrightarrow \mathcal{P}(L) \times \mathcal{P}(M) \times \mathcal{M}(\{I\} \times (\mathcal{L} \cup \mathcal{M})) \times \mathcal{P}(\{I\}) \]

\[ \Sigma \leftrightarrow (\Sigma_L, \Sigma_M, \Sigma_{[I]}^{-1}, \Sigma_{[I]}), \]

and

\[ \text{tr}(\Sigma^{-1} x x^t) = \text{tr}(\Sigma_L^{-1} x_L x_L^t) + \text{tr}(\Sigma_M^{-1} x_M x_M^t) + \text{tr}(\Sigma_{[I]}^{-1} (x_{[I]} - \Sigma_{[I]}^{-1} x_L x_L^t) (\cdots)^t). \]

Now \(L, M, I\) is a never-decreasing listing of \(J(\mathcal{X})\), so \(\Sigma\) may be reconstructed from its ordered nontrivial \(\mathcal{X}\)-parameters \(\Lambda[L], \Lambda[M], R[I]\), \(\Lambda[I]\) as follows:

**Step 1,2:** 
Repeat Steps 1,2 in Example 2.3.

**Step 3:**
\[ \Sigma_{[I]} = R[I] \text{Diag}(\Sigma_L, \Sigma_M) \]
\[ \Sigma_{[I]} = \Lambda[I] + R[I] \Sigma_{[I]} \]

Thus \(\mathcal{P}_I(I)\) consists of all \(\Sigma\) of the form

(2.57) \[ \Sigma = \begin{bmatrix} \Sigma_L & 0 & \cdots & \cdots & \cdots \\ 0 & \Sigma_M & \cdots & \cdots & \cdots \\ \vdots & \vdots & \ddots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \ddots & \cdots \\ \Sigma_{[I]} & x_{[I]} & \cdots & \cdots & \cdots \end{bmatrix} \]

where \(\Sigma\) is partitioned according to the ordered decomposition

(2.58) \[ I = L \cup M \cup [I]. \]

The precision matrix \(\Lambda \equiv \Sigma^{-1} \in \mathcal{P}_I(I)^{-1}\) is characterized by the condition that \(\Sigma_{[I]}^{-1} = \text{Diag}(\Sigma_L^{-1}, \Sigma_M^{-1})\). Thus, unlike the preceding example, here \(\mathcal{P}_I(I)^{-1}\) is linear while \(\mathcal{P}_I(I)^{-1}\) is not. The group \(\text{GL}_I(I)\) consists of all nonsingular \(I \times I\) matrices of the form
(2.59) \[ A = \begin{bmatrix} A_L & 0 & \cdots & 0 \\ 0 & A_M & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_I : A_I \end{bmatrix} \]

Example 2.5. Suppose that \( \mathcal{A} = \{ \emptyset, L \cap M, L, M, L \cup M \equiv I \} \) (see Figure 2.5). 
(Note that (1.3) is a special case.)

![Figure 2.5.](image)

Now \( J(\mathcal{A}) = \{ L \cap M, L, M \} \), and \( \langle L \cap M \rangle = \emptyset, \langle L \rangle = \langle M \rangle = L \cap M \). The \( \mathcal{A} \)-parametrization of \( P_\mathcal{A}(I) \) is given by

\[
(2.60) \quad P_\mathcal{A}(I) \leftrightarrow P(L \cap M) \times P([L] \times (L \cap M)) \times P([L]) \times P([M] \times (L \cap M)) \times P([M])
\]

\[ \Sigma \leftrightarrow (\Sigma_{L \cap M}, \Sigma_{[L]} \Sigma_{L \cap M}^{-1}, \Sigma_{[L]}, \Sigma_{[M]} \Sigma_{L \cap M}^{-1}, \Sigma_{[M]}). \]

and

\[
(2.61) \quad \text{tr}(\Sigma^{-1}xx^t) = \text{tr}(\Sigma_{L \cap M}^{-1}xx^t) \\
\quad + \text{tr}(\Sigma_{[L]}^{-1}(x[L]) - \Sigma_{[L]} \Sigma_{L \cap M}^{-1}x(L \cap M)(\cdots)^t) \\
\quad + \text{tr}(\Sigma_{[M]}^{-1}(x[M]) - \Sigma_{[M]} \Sigma_{L \cap M}^{-1}x(L \cap M)(\cdots)^t).
\]

Since \( L \cap M, L, M \) is a never-decreasing listing of \( J(\mathcal{A}) \), \( \Sigma \) may be reconstructed from its ordered \( \mathcal{A} \)-parameters \( \Lambda_{[L \cap M]}, R_{[L]}, \Lambda_{[L]}, R_{[M]} \), \( \Lambda_{[M]} \) as follows:

**Step 1:** \[ \Sigma_{L \cap M} = \Lambda_{[L \cap M]} \]
Step 2:  
\[ \Sigma_{[L]} = R_{[L]} \Sigma_{\{L\}} \]
\[ \Sigma_{[L]} = \Lambda_{[L]} + R_{[L]} \Sigma_{\{L\}} \]

Step 3:  
\[ \Sigma_{[M]} = R_{[M]} \Sigma_{\{L,M\}} \]
\[ \Sigma_{[M]} = \Lambda_{[M]} + R_{[M]} \Sigma_{\{L,M\}} \]
\[ \Sigma_{[M]} = R_{[M]} \Sigma_{\{L,M\}} \]
\[ \Sigma_{[M]} = \Lambda_{[M]} + R_{[M]} \Sigma_{\{L,M\}} \]
\[ \Sigma_{[M]} = \Lambda_{[M]} + R_{[M]} \Sigma_{\{L,M\}} \]

(2.62)

(Note that \( \Sigma_{[M]} = \Sigma_{\{L\}} \).) Thus \( P_\mathcal{M}(I) \) consists of all \( \Sigma \in P(I) \) of the form

\[ \Sigma = \begin{bmatrix}
\Sigma_{\{L,M\}} & \Sigma_{\{M\}} \\
\Sigma_{[L]} & \Sigma_{\{L\}} \\
\Sigma_{[M]} & \Sigma_{\{M\}}
\end{bmatrix} \]

(2.63)  
such that \( \Sigma_{[M]} \) satisfies (2.62) and where \( \Sigma \) is partitioned according to the ordered decomposition

\[ I = (\{L,M\}) \cup [L] \cup [M]. \]

Then \( P_\mathcal{M}(I)^{-1} \) consists of all \( \Lambda \in P(I) \) having the simple form

\[ \Lambda = \begin{bmatrix}
\Lambda_{\{L,M\}} & \Lambda_{\{M\}} \\
\Lambda_{\{L\}} & 0 \\
\Lambda_{\{M\}} & 0 & \Lambda_{[M]}
\end{bmatrix} \]

(2.65)

Thus, in this example \( P_\mathcal{M}(I)^{-1} \) is linear while \( P_\mathcal{M}(I) \) is not. The group \( GL_\mathcal{M}(I) \) consists of all nonsingular \( 1 \times 1 \) matrices of the form

\[ A = \begin{bmatrix}
\Lambda_{\{L,M\}} & 0 & 0 \\
\Lambda_{\{L\}} & 0 & 0 \\
\Lambda_{\{M\}} & 0 & \Lambda_{[M]}
\end{bmatrix} \]

(2.66)
Example 2.6. Consider the lattice $\mathcal{X} = \{\emptyset, \bigwedge, L, M, \bigvee, I\}$ (see Figure 2.6).

Figure 2.6.

Note that $J(\mathcal{X}) = \{\emptyset, \bigwedge, L, M, I\}$ and $\langle \bigwedge \rangle = \emptyset$, $\langle L \rangle = \langle M \rangle = \bigwedge$, $\langle I \rangle = \bigvee$. The $\mathcal{X}$-parametrization of $P_\mathcal{X}(I)$ is given by

\begin{equation}
(2.67)
\begin{align*}
P_\mathcal{X}(I) &\leftrightarrow \\
P(\bigwedge) \times P([L]) \times P([\bigwedge]) \times P([M]) \times P([\bigwedge]) \times P([I])
\end{align*}
\end{equation}

\begin{equation}
(2.68)
\begin{align*}
\Sigma &\leftrightarrow \\
&= (\Sigma_\bigwedge, \Sigma_{[L]}^{\bigwedge}, \Sigma_{[L]}^{\bigwedge}, \Sigma_{[M]}^{\bigwedge}, \Sigma_{[I]}^{\bigwedge}, \Sigma_{[I]}^{\bigwedge}).
\end{align*}
\end{equation}

Since $\bigwedge, L, M, I$ is a never-decreasing listing of $J(\mathcal{X})$, $\Sigma$ can be reconstructed from $A_{[\bigwedge]}$, $R_{[L]}$, $A_{[L]}$, $R_{[M]}$, $A_{[M]}$, $R_{[I]}$, $A_{[I]}$ as follows:

Steps 1, 2, 3: Repeat Steps 1, 2 and 3 in Example 2.5 to obtain $\Sigma_{\bigwedge}$. 

Step 4: 

\begin{align*}
\Sigma_{[I]} &= R_{[I]}^{\Sigma_{\bigwedge}} \\
\Sigma_{[I]} &= A_{[I]} + R_{[I]}^{\Sigma_{[I]}}.
\end{align*}
Thus \( P_\mathcal{X}(I) \) consists of all \( \Sigma \) of the form

\[
\Sigma = \begin{bmatrix}
\Sigma_{\mathcal{L} \mathcal{M}} & \Sigma_{<I>}
\end{bmatrix}
\]

(2.69)

partitioned according to \( I = I_{\mathcal{L} \mathcal{M}} \cup D_I \), where \( \Sigma_{\mathcal{L} \mathcal{M}} \) is given by (2.63), (2.64) and (2.62). The precision matrix \( \Lambda = \Sigma^{-1} \in P_\mathcal{X}(I)^{-1} \) is characterized by the condition that \( \Sigma^{-1}_{\mathcal{L} \mathcal{M}} \) have the form (2.65). Thus neither \( P_\mathcal{X}(I) \) nor \( P_\mathcal{X}(I)^{-1} \) is linear. The group \( \text{GL}_\mathcal{X}(I) \) consists of all nonsingular \( I \times I \) matrices of the form

\[
\Lambda = \begin{bmatrix}
A_{\mathcal{L} \mathcal{M}} & 0 & 0 & \vdots & 0 \\
A_{[L]} & A_{[L]} & 0 & \vdots & 0 \\
A_{[M]} & 0 & A_{[M]} & \ddots & 0 \\
& & & \ddots & \ddots \\
A_{[I]} & \vdots & \vdots & \ddots & A_{[I]}
\end{bmatrix}
\]

(2.70)

Example 2.7. Let \( \mathcal{X} \) be the lattice in Figure 2.7:

![Figure 2.7](image)

Then \( J(\mathcal{X}) = \{\mathcal{L} \mathcal{M}, \mathcal{L}, \mathcal{M}, \mathcal{L}', \mathcal{M}'\} \) and \( \langle \mathcal{L} \mathcal{M} \rangle = \emptyset, \langle \mathcal{L} \rangle = \langle \mathcal{M} \rangle = \mathcal{L} \mathcal{M}, \langle \mathcal{L}' \rangle = \langle \mathcal{M}' \rangle = \mathcal{L} \mathcal{M} \equiv \mathcal{L}' \mathcal{M}' \). The \( \mathcal{X} \)-parametrization of \( P_\mathcal{X}(I) \) is given by
from which the decomposition of \( \text{tr}(\Sigma^{-1}xx^t) \) is directly obtained. The matrix can be reconstructed from its ordered parameters \( \Lambda_{[L'M]} \), \( R_{[L']} \), \( \Lambda_{[L]} \), \( R_{[M']} \), \( \Lambda_{[L']}, R_{[M']} \), \( \Lambda_{[M']} \) as follows:

**Steps 1, 2, 3:** Repeat Steps 1, 2, 3 in Example 2.5 to obtain \( \Sigma_{L'M} = \Sigma_{L' \cap M'} \).

**Steps 4, 5:** Repeat Steps 2, 3 in Example 2.5 with \( L, M \) replaced by \( L', M' \).

Thus \( P_{\mathcal{A}}(I) \) consists of all \( \Sigma \) of the form (2.63) with \( L, M \) replaced by \( L', M' \), partitioned according to the ordered decomposition

\[
I = (L'M') \cup [L'] \cup [M']
\]

and where \( \Sigma_{L' \cap M'} \equiv \Sigma_{L'M} \) is given by (2.63). The precision matrix \( \Delta \equiv \Sigma^{-1} \) has the form (2.65) with \( L, M \) replaced by \( L', M' \) and satisfies the condition that \( \Sigma_{L'M}^{-1} \) has the form (2.65). Again, neither \( P_{\mathcal{A}}(I) \) nor \( P_{\mathcal{A}}(I)^{-1} \) is linear. The group \( \text{GL}_{\mathcal{A}}(I) \) consists of all nonsingular \( I \times I \) matrices of the form

\[
(2.71) \quad P_{\mathcal{A}}(I) \leftrightarrow \\
\Sigma \leftrightarrow \\
(\Sigma_{L'M} \cdot \Sigma_{[L]} \Sigma_{L'M}^{-1} \cdot \Sigma_{[L]} \cdot \Sigma_{[M']} \cdot \Sigma_{[M']}) \\
\Sigma_{[L']} \Sigma_{L'M}^{-1} \cdot \Sigma_{[L']} \cdot \Sigma_{[M']} \cdot \Sigma_{[M']}.
\]
Example 2.8. Let \( \mathfrak{X} \) be the lattice in Figure 2.8:

![Figure 2.8](image)

Here \( J(\mathfrak{X}) = \{L \cap M, L, M, L'\} \) and \( \langle L \cap M \rangle = \emptyset, \langle L \rangle = \langle L' \rangle = L \cap M, \langle L'' \rangle = L, \langle M' \rangle = L \cup M = L' \cap M' \). The \( \mathfrak{X} \)-parametrization of \( P_\mathfrak{X}(I) \) is given by:

\[
P(\mathfrak{X}) = P(\mathfrak{X}) \leftrightarrow 
\]

\[
\begin{align*}
\Sigma &\to \\
(\Sigma_{L \cap M}, \Sigma_{L}, \Sigma_{L'}, \Sigma_{L''}, \Sigma_{M}, \Sigma_{M'}) &\to (\Sigma_{L \cap M}, \Sigma_{L}, \Sigma_{L'}, \Sigma_{L''}, \Sigma_{M}, \Sigma_{M'})
\end{align*}
\]

from which the decomposition of \( \text{tr}(\Sigma^{-1} \times \Sigma) \) follows directly. The matrix \( \Sigma \) can be reconstructed from its ordered \( \mathfrak{X} \)-parameters \( \Lambda_{L \cap M}, R_{L}, \Lambda_{L'}, R_{L''}, \Lambda_{M}, R_{M'} \) as follows:

\[
(2.73) \quad \Lambda = \begin{bmatrix}
\Lambda_{L \cap M} & 0 & 0 & 0 & 0 \\
0 & \Lambda_{L} & 0 & 0 & 0 \\
0 & 0 & \Lambda_{L'} & 0 & 0 \\
A_{M} & 0 & A_{M'} & 0 & 0 \\
A_{M'} & 0 & A_{M'} & 0 & 0 \\
\end{bmatrix}
\]

(2.74) \quad P(\mathfrak{X}) = P(\mathfrak{X}) \leftrightarrow 

\[
\begin{align*}
P(L \cap M) &\times P([L] \times (L \cap M)) \times P([L]) \times P([M] \times (L \cap M)) \\
\times P([L']) \times P([L'']) \times P([M'] \times (L \cup M)) \times P([M'])
\end{align*}
\]

\[
\Sigma &\to \\
(\Sigma_{L \cap M}, \Sigma_{L}, \Sigma_{L'}, \Sigma_{L''}, \Sigma_{M}, \Sigma_{M'}) &\to (\Sigma_{L \cap M}, \Sigma_{L}, \Sigma_{L'}, \Sigma_{L''}, \Sigma_{M}, \Sigma_{M'})
\]
Steps 1, 2, 3: Repeat Steps 1, 2, 3 in Example 2.5, to obtain \( \Sigma_{L \cup M} \equiv \Sigma_{L \cap M'} \).

Step 4:

\[
\Sigma_{[L']} = R_{[[L']]} \Sigma_L
\]
\[
\Sigma_{[L']} = A_{[[L']]} + R_{[[L']]} \Sigma_{[L']}
\]

\[ (2.75) \]

\[
(\Sigma_{[L']} \Sigma_{[L']}^{-1}) \Sigma_{[M]} = (\Sigma_{[L']} \Sigma_{[L']}^{-1}) \Sigma_{[M']}
\]

where \( \Sigma_{\{M\}} = \Sigma_{\{M\}^\top} \); thus we obtain \( \Sigma_{L'} \).

Step 5:

\[
\Sigma_{[M']} = R_{[[M']]} \Sigma_{L \cup M}
\]
\[
\Sigma_{[M']} = A_{[[M']]} + R_{[[M']]} \Sigma_{[M']}
\]

\[ (2.76) \]

\[
(\Sigma_{[M']} \Sigma_{[M']}^{-1}) \Sigma_{[L']} = (\Sigma_{[M']} \Sigma_{[M']}^{-1}) \Sigma_{[L']}
\]

where \( \Sigma_{\{L'\}} = \Sigma_{\{L'\}^\top} \).

Thus \( P_{\Lambda}(I) \) consists of all \( \Sigma \) of the form

\[ (2.77) \]

\[
\Sigma = \\
\begin{pmatrix}
\Sigma_{[L \cup M]} & \Sigma_{[L]} & \Sigma_{[L']} & \Sigma_{[M]} & \Sigma_{[M']}
\end{pmatrix}
\]

partitioned according to the ordered decomposition

\[ (2.78) \]

\[
I = (L \cup M) \cup [L] \cup [M] \cup [L'] \cup [M'].
\]
where $\Sigma_{[M]}$, $\Sigma_{[L]}$, $\Sigma_{[M']}$ satisfy (2.62), (2.75), (2.76), respectively.

The precision matrix $\Lambda = \Sigma^{-1}$ satisfies the following three conditions:

- Its $[M'] \times [L']$- and $[L'] \times [M']$-submatrices are 0.
- The $[L'] \times [M]$- and $[M] \times [L']$-submatrices of $\Sigma^{-1}$ are 0.
- $\Sigma^{-1}_{L \cup M}$ has the form (2.65). Neither $P_{\mathcal{A}}(I)$ nor $P_{\mathcal{A}}(I)^{-1}$ is linear. The group $\mathcal{GL}_{\mathcal{A}}(I)$ consists of all nonsingular $I \times I$ matrices of the form

\[
(2.79) \quad \Lambda = \begin{bmatrix}
A_{L \cap M} & 0 & 0 & 0 & 0 \\
A_{[L]} & A_{[M]} & 0 & 0 & 0 \\
A_{[M]} & 0 & A_{[L']} & 0 & 0 \\
A_{[L']} & 0 & A_{[L']} & 0 & 0 \\
A_{[M']} & 0 & A_{[M']} & 0 & 0
\end{bmatrix}.
\]

\[\square\]

**Example 2.9.** Finally consider the lattice $\mathcal{A}$ in Figure 2.9a:

![Figure 2.9a](image)

The lattice $\mathcal{A}$.

Although this lattice properly contains the lattices in Examples 2.7 and 2.8 as sublattices, the set $P_{\mathcal{A}}(I)$ that it determines is much simpler than those in Examples 2.7 and 2.8. The reader may verify that $P_{\mathcal{A}}(I)$ is identical to $P_{\mathcal{M}}(I)$, where $\mathcal{M}$ is the sublattice in Figure 2.9b:
Remark 2.5. For any $K \in \mathcal{D}(I)$ define $K' := I \setminus K$. It is an elementary exercise to verify that for $L, M \in \mathcal{D}(I)$, $x_L \parallel x_M \mid x_{L \cap M}$ under $N(\Sigma)$ if and only if $x_L, \parallel x_M \mid x_{L \cap M}$ under $N(\Sigma^{-1})$. From this it follows that $P_\mathcal{X}(I) = P_{\mathcal{X}'}(I)^{-1}$, where $\mathcal{X}' := \{K' \mid K \in \mathcal{X}\}$ is the dual lattice of $\mathcal{X}$. For example, if $\mathcal{X}$ is the lattice in Figure 2.4, then $\mathcal{X}'$ has the same form as the lattice in Figure 2.5; the relation $P_\mathcal{X}(I) = P_{\mathcal{X}'}(I)^{-1}$ may be verified by comparing (2.57) and (2.65).

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

3.1. Factorization of the likelihood function: the MLE of $\Sigma$.

Consider $n$ independent, identically distributed (i.i.d.) observations $x_1, \ldots, x_n$ from the lattice CI model $N(\mathcal{X})$ defined by (1.8) and (1.4), and denote the matrix of observations by $y$, i.e.,

$$y := (x_1, \ldots, x_n) \in \mathcal{M}(I \times N).$$

where $N = \{1, \ldots, n\}$. For $L \in \mathcal{X}$ let $y_L$ denote the $L \times N$ submatrix of $y$.

while for $K \in J(\mathcal{X})$ partition $y_K$ according to (2.3) as follows:
The fundamental factorization of the LF for the model \( \mathcal{M}(\mathfrak{A}) \) is an immediate consequence of Theorem 2.1(ii), Lemma 2.5, and Theorem 2.2.

**Theorem 3.1.** (Factorization Theorem.) The likelihood function based on \( n \) i. i. d. observations from the statistical model \( \mathcal{M}(\mathfrak{A}) \) has the following factorization:

\[
\begin{align*}
P_{A}(I) \times M(I \times N) & \to ]0, \infty[ \\
(\Sigma, y) & \to (\det(\Sigma))^{-n/2} \exp(-\text{tr}(\Sigma^{-1}yy^t)/2) = \\
\Pi(\det(\Sigma[K]))^{-n/2} \exp(\text{tr}(\Sigma^{-1}y[K],(y[K] - \Sigma[K]^{-1}y[\mathcal{K}]y[K]^{-1}y[K])^t)/2) & |K \in J(\mathfrak{A})|.
\end{align*}
\]

The parameter space \( P_{A}(I) \) has the factorization given by (2.32). □

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

It follows readily from Theorem 3.1 and well-known results for the multivariate normal linear regression model that the MLE \( \hat{\Sigma}(y) \) of \( \Sigma \in P_{A}(I) \) is unique if it exists, and it exists for a.e. \( y \in M(I \times N) \) if and only if

\[
(3.3) \quad n \geq \max\{ |\mathcal{K}| + |[K]| \ | K \in J(\mathfrak{A}) \} \equiv \max\{ |K| \ | K \in J(\mathfrak{A}) \}.
\]

In this case the \( \mathfrak{A} \)-parameters of \( \hat{\Sigma} \) are determined from the usual formulas for regression estimators:
(3.4) \[ \sum_{[K]} S^{-1} = S_{[K]} S^{-1} \] \[ n_{[K]} = S_{[K]} \] \[ K \in J(\mathcal{X}) \]

where \( S(y) = yy^t \) is the empirical covariance matrix. The explicit expression for \( s^\Delta \) itself may be obtained from its \( \mathcal{X} \)-parameters in (3.4) by means of the reconstruction algorithm given in Section 2.7.

If \( I \in J(\mathcal{X}) \) then the condition (3.3) reduces to \( N \geq |I| \), so in this case \( S \) is positive definite a.e., hence a fortiori \( s^\Delta \) exists a.e. for every \( K \in J(\mathcal{X}) \). If, on the other hand, \( I \notin J(\mathcal{X}) \), then condition (3.3) does not guarantee that \( S \) is positive definite, but it still guarantees that \( s^\Delta \) (and hence \( s^\Delta \)) exists a.e.

By Lemma 2.5, when (3.3) is satisfied the maximum value of the LF (3.2) is given by

\[ c \cdot \Pi((\det(s^\Delta_{[K]}))^n/2 | K \in J(\mathcal{X})) = c \cdot (\det(s^\Delta))^n/2 \]

where \( c = n^{n/2} \exp(-n|I|/2) \). This fact is used in Section 4 to express the likelihood ratio statistic for testing one model against another.

Remark 3.1. The statistical model \( \mathcal{M}(\mathcal{X}) \) is a curved exponential family; it is linear if and only if \( P_{\mathcal{X}}(I)^{-1} \) is a linear set, i.e., closed under positive linear combinations. In the linear case the MLE \( \hat{\Delta} \) based on n i.i.d. observations from \( \mathcal{M}(\mathcal{X}) \) is a minimal sufficient statistic, but \( \hat{\Delta} \) is not necessarily sufficient in the general case.

3.2. Examples of pairwise conditional independence models.

For each lattice \( \mathcal{X} \) in Examples 2.1-2.9, consider the associated normal model \( \mathcal{N}(\mathcal{X}) \). When \( \mathcal{X} \) is a chain as in Examples 2.1 and 2.2, \( P_{\mathcal{X}}(I) = P(I) \)
and \( \mathcal{N}(\mathbf{x}) \) is the unrestricted covariance model regardless of the length of the chain. (The \( \mathcal{X} \)-parametrization of \( P_{\mathcal{X}}(I) \) does depend on this length, however.) Condition (3.3) for existence of the MLE \( \hat{\mathbf{A}} \) reduces to the familiar condition \( n \geq |I| \), while (3.4) reduces to \( n_A = S \).

For the lattice \( \mathcal{X} \) in Example 2.3, partition the observation \( \mathbf{x} \in \mathbb{R}^I \) according to (2.54) as \( \mathbf{x} = (x_L^t, x_M^t)^t \). The model \( \mathcal{N}(\mathcal{X}) \) states simply that \( x_L \parallel x_M \). According to (3.3), the MLE \( \hat{\mathbf{A}} \) exists if and only if \( n \geq \max\{|L|, |M|\} \) (whereas \( S \) is positive definite if and only if \( n \geq |I| \)) and is given by \( n_A = \text{Diag}(S_L, S_M) \).

For the lattice \( \mathcal{X} \) in Example 2.4, partition \( \mathbf{x} \in \mathbb{R}^I \) according to (2.58) as \( \mathbf{x} = (x_L^t, x_M^t, x_{[I]}^t)^t \). Then the model \( \mathcal{N}(\mathcal{X}) \) again states that \( x_L \parallel x_M \). Condition (3.3) for the existence of the MLE takes the form \( n \geq |I| \), while from (3.4).

\[
\begin{align*}
n_A^{\mathcal{L}} &= S_L, \\
n_A^{\mathcal{M}} &= S_M, \\
\Sigma_{[I]}^{-1} &= S_{[I]}^{-1} S_{\mathcal{L}} S_{\mathcal{M}}, \\
n_A^{[I]} &= S_{[I]}.
\end{align*}
\]

We reconstruct \( \hat{\mathbf{A}} \) from its \( \mathcal{X} \)-parameters by following Steps 1-3 in Example 2.4 to obtain

\[
\begin{align*}
n_A^{\mathcal{LUM}} &= \text{Diag}(S_L, S_M) \\
n_A^{[I]} &= S_{[I]}^{-1} S_{[I]} \text{Diag}(S_L, S_M) \\
\hat{n}_A^{[I]} &= S_{[I]}, + S_{[I]} (\text{Diag}(S_L, S_M))^{-1} S_{[I]} (\neq S_{[I]}).
\end{align*}
\]

In Example 2.5, \( x \) is partitioned according to (2.64) as \( (x_L^t, x_M^t, x_{[I]}^t)^t \). The model \( \mathcal{N}(\mathcal{X}) \) states that \( x_{[L]} \parallel x_{[M]}; x_{L\cap M} \).

Condition (3.3) becomes \( n \geq \max\{|L|, |M|\} \), while (3.4) becomes
By Steps 1-3 in Example 2.5, \( \mathbf{A} \) is given by (3.6a) and

\[(3.6a) \quad n^A_{[\mathbf{L} \mathbf{M}]} = S_{\mathbf{L} \mathbf{M}}\]
\[(3.6b) \quad \Sigma_{[\mathbf{L} \mathbf{M}]}^{-1} = S_{[\mathbf{L}]S_{\mathbf{L} \mathbf{M}}}, \quad n^A_{[\mathbf{L}]} = S_{[\mathbf{L}]}, \]
\[(3.6c) \quad \Sigma_{[\mathbf{M} \mathbf{L} \mathbf{M}]}^{-1} = S_{[\mathbf{M}]S_{\mathbf{L} \mathbf{M}}}, \quad n^A_{[\mathbf{M}]} = S_{[\mathbf{M}]}.\]

In Example 2.6, \( x \) is partitioned as \( (x^t_{\mathbf{L} \mathbf{M}}, x^t_{[\mathbf{L}]}, x^t_{[\mathbf{M}]}, x^t_{[\mathbf{I}]} \)^t \) and the model \( N(x) \) states that \( x_{[\mathbf{L}]} \parallel x_{[\mathbf{M}]} | x_{\mathbf{L} \mathbf{M}} \). Condition (3.3) reduces to \( n \geq |I| \), while (3.4) is given by (3.6a,b,c) and

\[(3.7a) \quad n^A_{[\mathbf{L}]} = S_{[\mathbf{L}]}, \quad n^A_{[\mathbf{L}]} = S_{[\mathbf{L}]}
\[(3.7b) \quad n^A_{[\mathbf{M}]} = S_{[\mathbf{M}]}, \quad n^A_{[\mathbf{M}]} = S_{[\mathbf{M}]}
\[(3.7c) \quad n^A_{[\mathbf{M}]} = S_{[\mathbf{M}]S_{\mathbf{L} \mathbf{M}}S_{[\mathbf{L}]}^\top} (\neq S_{[\mathbf{M}]}) \].

From Steps 1-4 in Example 2.6, \( \mathbf{A} \) is given by (3.6a), (3.7a,b,c), and

\[(3.8) \quad n^A_{\mathbf{L} \mathbf{M}} = \begin{bmatrix} S_{\mathbf{L} \mathbf{M}} & S_{[\mathbf{L}]} & S_{[\mathbf{M}]} \\ S_{[\mathbf{L}]} & S_{[\mathbf{L}]} & S_{[\mathbf{L}]S_{\mathbf{L} \mathbf{M}}S_{[\mathbf{M}]}^\top} \\ S_{[\mathbf{M}]} & S_{[\mathbf{M}]S_{\mathbf{L} \mathbf{M}}S_{[\mathbf{L}]}^\top} & S_{[\mathbf{M}]} \end{bmatrix}.\]
In Example 2.7, \( x \) is partitioned as \( (x_{L}^{t} x_{[L]}^{t} x_{[M]}^{t} x_{[L']}^{t} x_{[M']}^{t})^{t} \) and the model \( N(\mathcal{M}) \) states that

\[
x_{[L]} \perp x_{[M]} | x_{L \cap M} \quad \text{and that} \quad x_{[L']} \perp x_{[M']} | (x_{L \cap M}, x_{[L]}, x_{[M]}).
\]

(Note that \( x_{L \cap M} = x_{L \cup M} = (x_{L \cap M} x_{[L]} x_{[M]}) \).) Condition (3.3) becomes \( n \geq \max\{|L'|, |M'|\} \), while (3.4) is given by (3.6a,b,c) and (3.6b,c) with \( L, M \) replaced by \( L', M' \) (note that \( S_{L \cap M'} = S_{L \cup M} \)). From Steps 1-5 in Example 2.7, \( \Delta \) is given by (3.6a), (3.7a,b,c), and

\[
\begin{align*}
\Delta_{[L']}^{A} &= S_{[L']} S_{L \cup M}^{-1} (n_{L \cup M}^{A}) \\
\Delta_{[L]}^{A} &= S_{[L]} + S_{[L']} S_{L \cup M}^{-1} (n_{L \cup M}^{A}) S_{L \cup M}^{-1} S_{[L']}
\end{align*}
\]

(3.9a)

\[
\begin{align*}
\Delta_{[M']}^{A} &= S_{[M']} S_{M \cup M}^{-1} (n_{M \cup M}^{A}) \\
\Delta_{[M]}^{A} &= S_{[M]} + S_{[M']} S_{M \cup M}^{-1} (n_{M \cup M}^{A}) S_{M \cup M}^{-1} S_{[M']}
\end{align*}
\]

(3.9b)

(3.9c)

where \( n_{L \cup M}^{A} \) is given by (3.8).

In Example 2.8, \( x \) is partitioned as \( (x_{L}^{t} x_{[L]}^{t} x_{[M]}^{t} x_{[L'']}^{t} x_{[M'']}^{t})^{t} \). It may be seen from the form (2.77) of \( \Sigma \in P_{\mathcal{A}}(I) \) that the model \( N(\mathcal{M}) \) is determined by the following three conditions:

\[
\begin{align*}
(1) \quad & x_{[L]} \perp x_{[M]} | x_{L \cap M} \\
(2) \quad & x_{[M]} \perp x_{[L'']} | (x_{L \cap M}, x_{[L]}) \\
(3) \quad & x_{[L'']} \perp x_{[M']} | (x_{L \cap M}, x_{[L]}, x_{[M]}).
\end{align*}
\]

Condition (3.3) becomes \( n \geq \max\{|L'\prime|, |M'\prime|\} \), while (3.4) is given by (3.6a,b,c).
From Steps 1-5 in Example 2.8, $\xi$ is given by (3.9a) and (3.7a,b,c), by

$$
\xi^{\xi}_{[L^0]} = S_{[L^0]}^{S_{[L^0]}^{-1}}, \quad n^{\xi}_{[L^0]} = S_{[L^0]}
$$

Finally, for the lattice $\mathcal{X}$ in Example 2.9, $x$ is partitioned as

$$(x_{L\cap M}, x_{[L]}, x_{[M]}, x_{[L^0]}, x_{[M^0]})^t.$$  It readily seen (cf. Remark 3.2) that the model $N(\mathcal{X})$ is determined by the single condition that

$$
(x_{[L]}, x_{[L^0]}) \parallel (x_{[M]}, x_{[M^0]}) \mid x_{L\cap M}.
$$

This reflects the fact that this model is of the same form as that in Example 2.5 (see the discussion in Example 2.9).

Remark 3.2. Recall the definition of the normal model $N(\mathcal{X})$ for a distributive lattice $\mathcal{X}$: for every pair $L, M \in \mathcal{X}$, $x_L \parallel x_M \mid x_{L\cap M}$. It may be seen from the above examples that many of these conditions are redundant and may be omitted, for example whenever $L \subseteq M$. More generally, if $L \subseteq$
L', M ⊆ M', and L ∩ M = L' ∩ M', then $x_L \perp x_M \mid x_{L' \cap M'} \Rightarrow x_L \perp x_M \mid x_{L \cup M}$, hence the latter condition may be omitted. The question of characterizing a minimal set of CI conditions that determines $\mathcal{N}(\mathfrak{X})$ is currently under investigation. For a given lattice $\mathfrak{X}$, however, such minimal determining sets are not unique. In Example 2.8, the following four sets of CI conditions are (equivalent) minimal determining sets for $\mathcal{N}(\mathfrak{X})$:

(i) $x_L \perp x_M \mid x_{L \cup M}$; (ii) $x_{L \cup M} \perp x_{L''} \mid x_L$; (iii) $x_L \perp x_M \mid x_{L \cup M}$; (iv) $x_{L''} \perp x_{L \cup M} \mid x_L$.

Remark 3.3. 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 $\mathcal{N}(\mathfrak{X})$ for any $\mathfrak{X}$. 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 3.4. The general model $\mathcal{N}(\mathfrak{X})$ is defined by the pairwise CI requirement (1.4) for every pair $L, M \in \mathfrak{X}$. This requirement does not necessary imply, however, that for every subset $\mathcal{Y} \subseteq \mathfrak{X}$, $(x_K|K \in \mathcal{Y})$ are mutually CI given $x_{\mathcal{Y} \cap (K|K \in \mathcal{Y})}$. For the lattice $\mathfrak{X}$ in Example 2.9, this may be seen by considering the subset $\mathcal{Y} = \{L'', L \cup M, M''\}$.

Remark 3.5. An alternative statistical interpretation of the CI model $\mathcal{N}(\mathfrak{X})$ may be obtained from (2.35): $x \equiv (x_K|K \in J(\mathfrak{X})) \in \mathbb{R}^I$ is an
observation from the normal model $N(\mathcal{M})$ if and only if $x$ can be represented in the form $x = Az$ for some (generalized block-triangular) matrix $A \in \mathcal{G}_\mathcal{M}(I)$, where $z \equiv (z_{[K]} | K \in J(\mathcal{M})) \in \mathbb{R}^I$ is an (unobservable) stochastic variate such that $z \sim N(1_I)$. From Proposition 2.2(iii), this representation is equivalent to the system of equations

$$
(3.10) \quad x[L] = \Sigma(A[L]z[M] | M \in H(L)), \quad L \in J(\mathcal{M}).
$$

where $H(L) := \{M \in J(\mathcal{M}) | M \subseteq L\} \equiv J(M_L)$. This shows that the CI model $N(\mathcal{M})$ 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 $(H(l) | l \in J)$ be a family of subsets of $J$ that satisfies the following two conditions:

(i) \quad $l \in H(l)$

(ii) \quad $m \in H(l) \Rightarrow H(m) \subseteq H(l)$.

For each $l \in J$ let $D_l$ and $E_l$ be finite index sets such that $|D_l| \leq |E_l|$ and let $I = \hat{U}(D_l | l \in J), I^* = \hat{U}(E_l | l \in J)$. Consider the normal statistical model defined by the system of equations

$$
(3.11) \quad x[l] = \Sigma(A_{lm}z[m] | m \in H(l)), \quad l \in J,
$$

where $x[l] \in \mathbb{R}^{D_l}$ is observable, $z[m] \in \mathbb{R}^{E_m}$ is unobservable, $z \equiv (z[m] | m \in J) \sim N(1_I)$ on $\mathbb{R}^I$, $A_{lm} \in \mathcal{M}(D_l \times E_m)$, and $\text{rank}(A_{ll}) = |D_l|$. Let $\mathcal{H}$ be the ring of subsets of $J$ generated by $\{H(l) | l \in J\}$ and for $H \in \mathcal{H}$ define $I_H = \hat{U}(D_l | l \in H)$. Then trivially $\mathcal{H} := \{I_H | H \in \mathcal{H}\}$ is a ring of
3.3. 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 $N(\mathscr{X})$ is invariant under the transitive action (2.33) of $\mathfrak{GL}(I)$ on the parameter space $\mathcal{P}_{\mathfrak{GL}}(I)$ and the action

$$\text{(3.12)} \quad \mathfrak{GL}(I) \times \mathcal{M}(I \times \mathbb{N}) \to \mathcal{M}(I \times \mathbb{N})$$

$$(A,y) \to Ay$$

of $\mathfrak{GL}(I)$ on the observation space $\mathcal{M}(I \times \mathbb{N})$. The MLE is thus equivariant.

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

Let $\mathfrak{X}$ and $\mathfrak{M}$ be two sublattices of $\mathfrak{D}(I)$ such that $\mathfrak{M} \subseteq \mathfrak{X}$. Then $\mathcal{P}_{\mathfrak{GL}}(I) \subseteq \mathcal{P}_{\mathfrak{GL}}(I)$ and one can consider the following general testing problem: based on $n$ i.i.d. observations $x_1, \ldots, x_n \in \mathbb{R}^I$ from the model $N(\mathfrak{M})$, test

$$\text{(4.1)} \quad H_0: \Sigma \in \mathcal{P}_{\mathfrak{GL}}(I) \quad \text{vs.} \quad H: \Sigma \in \mathcal{P}_{\mathfrak{GL}}(I).$$

In this section we find the likelihood ratio (LR) statistic $\lambda$ for this problem. The central distribution of $\lambda$, expressed in terms of its moments, is derived by means of the invariance of this testing problem under the actions of $\mathfrak{GL}(I)$ on the observation space and parameter space. This derivation is based on Theorem 4.2, which establishes the mutual independence of the maximal invariant statistic $\tau$ and the MLE's $\hat{\Sigma}^{[K]}$.
Examples of the general testing problem are presented in Section 4.3.

A warning about the notation is needed here. Since \( J(\mathcal{X}) \neq J(\mathcal{M}) \), quantities such as \( \langle K \rangle, [K], \Sigma_{[K]}, \Sigma_{[K]}, \Sigma_{[K]} \), depend not only on the subset \( K \) of \( I \) but also on the lattice of which \( K \) is considered a member. Thus, for example, \( \langle K \rangle_{\mathcal{X}} \) and \( \langle K \rangle_{\mathcal{M}} \) need not be the same. To alleviate this difficulty without introducing \( \mathcal{X} \) and \( \mathcal{M} \) as subscripts, the letter \( K \) shall denote a subset of \( I \) that is to be considered as a member of \( \mathcal{X} \), while \( M \) shall denote a subset of \( I \) that is to be considered a member of \( \mathcal{M} \).

4.1. The likelihood ratio statistic.

Denote the MLE's of \( \Sigma \) under \( \mathcal{N}(\mathcal{X}) \) and \( \mathcal{N}(\mathcal{M}) \) by \( \hat{\Sigma}_{\mathcal{X}} = \hat{\Sigma} \) and \( \hat{\Sigma}_{\mathcal{M}} = \hat{\Sigma} \), respectively.

**Theorem 4.1.** Suppose that \( n \geq \max\{|M| \mid M \in J(\mathcal{M})\} \). Then for every \( \Sigma \in \mathcal{P}(I) \), \( \hat{\Sigma} \) and \( \hat{\Sigma} \) exist a.e.. The LR statistic \( \lambda \) for testing \( H_0 \) against \( H \) is given by

\[
\lambda^{2/n} = \frac{\det(\hat{\Sigma})}{\det(\hat{\Sigma})} = \frac{\Pi(\det(\hat{\Sigma}_{\mathcal{M}})|M \in J(\mathcal{M}))}{\Pi(\det(\hat{\Sigma}_{[K]})|K \in J(\mathcal{X}))} = \frac{\Pi(\det(S_{[M]}_{\mathcal{M}})|M \in J(\mathcal{M}))}{\Pi(\det(S_{[K]}_{\mathcal{X}})|K \in J(\mathcal{X}))}.
\]

**Proof.** The first assertion follows from (3.3) and the inequality \( \max\{|M| \mid M \in J(\mathcal{M})\} \geq \max\{|K| \mid K \in J(\mathcal{X})\} \). To establish this inequality define the mapping \( \psi: J(\mathcal{X}) \to J(\mathcal{M}) \) by \( \psi(K) := \cap(M \in \mathcal{M} \mid M \supseteq K) \). By an argument similar to that in Proposition 3.2(ii) of [A] (1990) it may be shown that \( \psi \) is well-defined, while Proposition 3.3(i) of [A] (1990)
implies that $\psi$ is surjective, hence

\[
\max\{\left| M \right| \mid M \in J(\mathcal{M}) \} = \max\{\left| \psi(K) \right| \mid K \in J(\mathcal{X}) \} \\
\geq \max\{\left| K \right| \mid K \in J(\mathcal{X}) \}.
\]

The second assertion of the Theorem now follows from (3.5) and (3.4). \hfill \square

For computational purposes, note that

\[
(4.3) \quad \frac{\det(S_K(y))}{\det(S_{\langle K \rangle}(y))} = \frac{\det(y_{KK}^t)}{\det(y_{\langle K \rangle}^t)}. \tag{4.3}
\]

$K \in J(\mathcal{X})$, where $S(y) = yy^t$ (cf. (3.1)) with an analogous formula for $\det(S_{[M]}(y))$, $M \in J(\mathcal{M})$.

4.2. Central distribution and Box approximation.

The testing problem (4.1) is invariant under the action (3.12) of the group $\text{GL}_{\mathcal{M}}(I)$ on the sample space $\mathcal{M}(I \times N)$ and the action

\[
(4.4) \quad \text{GL}_{\mathcal{M}}(I) \times P_{\mathcal{M}}(I) \to P_{\mathcal{M}}(I) \\
(A, \Sigma) \to A\Sigma A^t
\]
on the parameter space. Let

\[
(4.5) \quad \pi: \mathcal{M}(I \times N) \to \mathcal{M}(I \times N)/\text{GL}_{\mathcal{M}}(I)
\]
denote the orbit projection (≡ maximal invariant) onto the orbit space.
under the action (3.12). Since the LR statistic is invariant under
(3.12), \( \lambda \) depends on \( y \in \mathbb{M}(I \times N) \) only through \( \pi(y) \). The central
distribution of \( \lambda \) is readily derived from this fact and Theorem 4.2,
whose proof is deferred to Appendix A.3. Since the restriction of (4.4)
to \( P_{\mathcal{A}}(I) \) is transitive (cf. Theorem 2.3), under \( H_0 \) the distribution of \( \lambda \)
does not depend on \( \Sigma \in P_{\mathcal{A}}(I) \).

**Theorem 4.2.** Under \( H_0 \), the statistics \( \pi \) and \( \hat{\lambda}_{[K]}, K \in J(\mathcal{A}) \), are mutually
independent. The statistic \( \hat{\lambda}_{[K]} \) has the Wishart distribution on \( P([K]) \)
with \( n-|K| \) degrees of freedom and expected value \( \Sigma_{[K]} \).

It follows from Theorem 4.2 that \( \lambda \) and \( \hat{\lambda}_{[K]}, K \in J(\mathcal{A}) \), are mutually
independent. Therefore for every \( \Sigma \in P_{\mathcal{A}}(I) \) (\( \subseteq P_{\mathcal{A}}(I) \)) and \( \alpha > 0 \),

\[
E((\det(\hat{\Sigma}))^\alpha) = E((\det(\hat{\Sigma}))^\alpha \lambda^{2\alpha/n}) = E((\det(\hat{\Sigma}))^\alpha)E(\lambda^{2\alpha/n}),
\]

hence from (2.37) and (4.2),

\[
E(\lambda^{2\alpha/n}) = \frac{E((\det(\hat{\Sigma}))^\alpha)}{E((\det(\hat{\Sigma}))^\alpha)} = \frac{\Pi(E((\det(\hat{\Sigma}_{[M]})^\alpha)|M \in J(\mathcal{A}))}{\Pi(E((\det(\hat{\lambda}_{[K]})^\alpha)|K \in J(\mathcal{A}))}
\]

However, it follows from the Wishart distribution of \( \hat{\lambda}_{[K]} \), that

\[
E((\det(\hat{\lambda}_{[K]})^\alpha) = (2n)^\alpha |[K]| (\det(\Sigma_{[K]})^\alpha \times \Pi \frac{\Gamma(n-|K|,-i+1)/2+\alpha)}{\Gamma((n-|K|,-i+1)/2)} |i=1,\ldots,|[K]|}
\]
for $K \in \mathcal{J}(\mathcal{X})$, with an analogous formula for $\mathbb{E}((\det(\tilde{X}_{[M]})^\alpha, M \in \mathcal{J}(\mathcal{A})$.

Since

$$\Sigma(\{[K] \mid K \in \mathcal{J}(\mathcal{X})\}) = |I| = \Sigma(\{[M] \mid M \in \mathcal{J}(\mathcal{A})\})$$

and

$$\Pi(\det(\Sigma_{[K]}^*) | K \in \mathcal{J}(\mathcal{X})) = \det(\Sigma) = \Pi(\det(\Sigma_{[M]}^*) | M \in \mathcal{J}(\mathcal{A}))$$

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

$$\mathbb{E}(\lambda^{2\alpha/n}) = \frac{\Pi \left[ \frac{\Gamma((n-|K|-j+1)/2+\alpha)}{\Gamma((n-|K|-j+1)/2)} \right]_{j=1,\ldots,|[K]|} | K \in \mathcal{J}(\mathcal{X})}{\Pi \left[ \frac{\Gamma((n-|M|-i+1)/2+\alpha)}{\Gamma((n-|M|-i+1)/2)} \right]_{i=1,\ldots,|[M]|} | M \in \mathcal{J}(\mathcal{A})}.$$ 

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

$$f = -2\sum((-|K|-j+1)/2 | j=1,\ldots,|[K]|) | K \in \mathcal{J}(\mathcal{X})$$

$$+2\sum((-|K|-j+1)/2 | j=1,\ldots,|[K]|) | K \in \mathcal{J}(\mathcal{X})$$

$$= \sum([K] | K \in \mathcal{J}(\mathcal{X}))$$

where the final equality is obtained using (4.6). From (2.37), one recognizes $f$ to be simply the usual difference between the number of free parameters under $H$ and the number of free parameters under $H_0$. 
4.3. Examples of testing problems.

Let $\mathcal{A}_1, \ldots, \mathcal{A}_9, \mathcal{A}_{10}, \mathcal{A}_{11}$ denote the lattices appearing in Figures 2.1, 2.8, 2.9a, 2.10a, 2.11a, respectively. In this subsection we consider examples of the testing problem (4.1) with $(\mathcal{A}, \mathcal{M}) = (\mathcal{A}_i, \mathcal{A}_j)$ for various pairs $(i, j)$. In each example the LR statistic $\lambda$ in (4.2) and the parameter $f$ in (4.8) is 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 $K, L \in P(I)$ such that $L \subseteq K$, 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 $\Sigma_K$ according to the decomposition

$$
K = L \cup (K \setminus L)
$$

and define

$$
\Sigma_{K \cdot L} = \Sigma_{K \setminus L} - \Sigma_{K \setminus L,L}^{-1} \Sigma_{L,K \setminus L} \in P(K \setminus L).
$$

(When $K \in J(\mathcal{A})$ and $M \in J(\mathcal{M})$, $\Sigma_K \cdot \langle K \rangle = \Sigma_{[K]}$, and $\Sigma_M \cdot \langle M \rangle = \Sigma_{[M]}$.) The well-known formula

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

may be applied in (4.2) to obtain the expressions for $\lambda^{2/n}$ that appear below.
First, set $\mathcal{M} = \{\emptyset, \{I\}\}$ in (4.1) and consider the testing problems of the form

\[(4.9) \quad H_0: \Sigma \in P_\mathcal{M}(I) \text{ vs. } H: \Sigma \in P(I)\]

for $\mathcal{M} = \mathcal{M}_3, \cdots, \mathcal{M}_8$. For $i = 3, \cdots, 8$, the following forms of the LR statistics $\lambda_i$ directly reflect the statistical interpretations of the models $\mathcal{N}(\mathcal{M}_i)$ given in Section 3.2:

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

$\lambda_3^{2/n} = \frac{\det(S)}{\det(S_L)\det(S_M)}$.

$f_3 = |L| \times |M|$;

$\mathcal{M} = \mathcal{M}_4$:

$\lambda_4^{2/n} = \frac{\det(S_{LM})}{\det(S_L)\det(S_M)}$.

$f_4 = |L| \times |M|$;

$\mathcal{M} = \mathcal{M}_5$:

$\lambda_5^{2/n} = \frac{\det(S_{I \setminus (L \cap M)})}{\det(S_L \cdot (L \cap M))\det(S_M \cdot (L \cap M))}$.

$f_5 = |[L]| \times |[M]|$;

$\mathcal{M} = \mathcal{M}_6$:

$\lambda_6^{2/n} = \frac{\det(S(L \cap M) \cdot (L \cap M))}{\det(S_L \cdot (L \cap M))\det(S_M \cdot (L \cap M))}$.

$f_5 = |[L]| \times |[M]|$;
\[ \lambda_7^{2/n} = \frac{\det(S_M \cdot (L \cap M))}{\det(S_L \cdot (L \cap M)) \det(S_M \cdot (L \cap M))} \times \frac{\det(S_I \cdot (L \cap M))}{\det(S_L \cdot (L \cap M)) \det(S_M \cdot (L \cap M))} \]

\[ f_7 = |L| \times |M| + |L'| \times |M'| \]

\[ \lambda_8^{2/n} = \frac{\det(S_M \cdot (L \cap M))}{\det(S_L \cdot (L \cap M)) \det(S_M \cdot (L \cap M))} \]

\[ \times \frac{\det(S_L \cdot L)}{\det(S_M \cdot (L \cap M)) \det(S_I \cdot (L \cap M))} \times \frac{\det(S_I \cdot (L \cap M))}{\det(S_L \cdot (L \cap M)) \det(S_M \cdot (L \cap M))} \]

\[ f_8 = |L| \times |M| + |M| \times |L'| + |L''| \times |M'| \]

\[ = |L| \times |M| + |L''| (|M| + |M'|) \]

\[ = |M| (|L| + |L''|) + |L''| \times |M'| \]

Remark 4.1. The three equivalent expressions for \( \lambda_8^{2/n} \) given above correspond to the first three minimal determining sets of CI conditions for \( N(M) \) given in Remark 3.2. The expression for \( \lambda_8^{2/n} \) suggested by the fourth set is

\[ \frac{\det(S_L \cdot (L \cap M))}{\det(S_M \cdot (L \cap M)) \det(S_I \cdot (L \cap M))} \times \frac{\det(S_I \cdot (L \cap M))}{\det(S_L \cdot (L \cap M)) \det(S_M \cdot (L \cap M))} \times \frac{\det(S_I \cdot (L \cap M))}{\det(S_L \cdot (L \cap M)) \det(S_M \cdot (L \cap M))} \]
but this is not equal to \( \lambda_8^{2/n} \). Thus the fourth determining set is in some sense unsatisfactory for describing \( \mathcal{N}(x_8) \).

Next we consider five testing problems of the form (4.1) with \((\mathcal{X}, \mathcal{M}) = (x_1, x_j)\). From (4.2) and (4.8) one may obtain the following expressions:

\[
(\mathcal{X}, \mathcal{M}) = (x_7, x_6): \quad \lambda_{7,6}^{2/n} = (\lambda_7/\lambda_6)^{2/n} = \frac{\det(S_{\text{L}} \cdot (LUM))}{\det(S_{\text{L}} \cdot (LUM)) \cdot \det(S_{\text{M}} \cdot (LUM))},
\]

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

\[
(\mathcal{X}, \mathcal{M}) = (x_8, x_7): \quad \lambda_{8,7}^{2/n} = (\lambda_8/\lambda_7)^{2/n} = \frac{\det(S_{\text{L}} \cdot (LUM))}{\det(S_{\text{L}} \cdot (LUM)) \cdot \det(S_{\text{M}} \cdot (LUM))},
\]

\[
f_{8,7} = f_8 - f_7 = |[M']| \times |[L']|;
\]

\[
(\mathcal{X}, \mathcal{M}) = (x_9, x_8): \quad \lambda_{9,8}^{2/n} = \frac{\det(S_{\text{M}} \cdot (M'))}{\det(S_{\text{M}} \cdot (M)) \cdot \det(S_{\text{M}} \cdot (M'))},
\]

\[
f_{9,8} = |[L]| \times |[M']|;
\]

\[
(\mathcal{X}, \mathcal{M}) = (x_{11}, x_6): \quad \lambda_{11,6}^{2/n} = \frac{\det(S_{\text{L}} \cdot (LUM))}{\det(S_{\text{L}} \cdot (LUM)) \cdot \det(S_{\text{L}} \cdot (LUM))},
\]

\[
f_{11,8} = |[M']| \times |[L']|;
\]

\[
(\mathcal{X}, \mathcal{M}) = (x_8, x_{11}): \quad \lambda_{8,11}^{2/n} = \lambda_{7,6}^{2/n}, \quad f_{8,11} = f_{7,6}.
\]
These five testing problems involve the five adjacent pairs of lattices in the diagram

\[ \mathcal{X}_6 \subset \mathcal{X}_7 \subset \mathcal{X}_8 \subset \mathcal{X}_9. \]

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

\[(\mathcal{X}_i, \mathcal{M}) = (\mathcal{X}_9, \mathcal{X}_7): \quad \lambda_{9,7}^{2/n} = (\lambda_{9,8} \cdot \lambda_{8,7})^{2/n}. \quad f_{9,7} = f_{9,8} + f_{8,7}.\]

Remark 4.2. 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 \( \lambda \) in (4.2) for the general testing problem (4.1) 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.

§5. INVARIANT FORMULATION OF THE CI MODEL AND TESTING PROBLEM.

5.1. The lattice structure of quotient spaces.

Let \( V \) be a finite-dimensional real vector space. A quotient space (or simply a quotient) of \( V \) is formally defined to be a pair \((Q, p_Q)\) consisting of a vector space \( Q \) and a surjective linear mapping \( p_Q: V \to Q \). For ease of notation, \((Q, p_Q)\) usually is abbreviated to \( Q \).
Let \( R \) and \( T \) be two quotients of \( V \). If there exists a linear mapping \( p_{RT}: T \to R \) such that \( p_R = p_{RT} \circ p_T \) then \( p_{RT} \) is necessarily surjective and unique, hence \( (R, p_{RT}) \) is a quotient of \( T \). In this situation we write \( (R, p_R) \leq (T, p_T) \), or simply \( R \leq T \). This relation is equivalent to the condition that \( p_R^{-1}(0) \supseteq p_T^{-1}(0) \). The relation \( \leq \) on the set of all quotients of \( V \) is not antisymmetric, hence one defines an equivalence relation \( \sim \) on this set by \( R \sim T \) if \( p_R^{-1}(0) = p_T^{-1}(0) \). The collection of equivalence classes is denoted by \( Q(V) \). Equipped with the relation induced by \( \leq \) (also denoted by \( \leq \)), \( Q(V) \) becomes a partially ordered set (\( \leq \) poset).

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

The poset \( Q(V) \) is in fact a lattice: if \( R, T \in Q(V) \) then their minimum and maximum exist and are given by

\[
R \land T := V/(p_R^{-1}(0) + p_T^{-1}(0))
\]
\[
R \lor T := V/(p_R^{-1}(0) \cap p_T^{-1}(0))
\]

respectively. The minimal and maximal elements exist and are given by \( \{0\} \) and \( V \) respectively. If \( \dim(V) \geq 2 \) then \( Q(V) \) is not distributive and \( |Q(V)| = \infty \). Since \( V \) is finite dimensional, the lattice \( Q(V) \) has finite length, hence so does any sublattice \( Q \subseteq Q(V) \). Therefore, if \( Q \) is a distributive sublattice of \( Q(V) \) it must be finite. The reader is referred to Section 3 of [A] (1990) for the properties of posets and lattices used here.
5.2. Invariant formulation of the pairwise CI model.

For \( \sigma \in \mathcal{P}(V) := \) the cone of all positive definite forms on the dual vector space \( V^* \) of \( V \), let \( N(\sigma) \) denote the normal distribution on \( V \) with mean vector \( 0 \in V \) and covariance \( \sigma \) (cf. [A] (1975), Section 5). Let \( Q \subseteq \mathcal{Q}(V) \) be a sublattice such that \( \{0\}, V \in Q \).

**Definition 5.1.** The class \( \mathcal{P}_Q(V) \subseteq \mathcal{P}(V) \) is defined as follows:

\[
\sigma \in \mathcal{P}_Q(V) \iff p_{R}(x) \parallel p_{T}(x) \mid p_{RAT}(x) \quad \forall R, T \in Q \text{ when } x \sim N(\sigma),
\]

i.e., \( p_{R} \) and \( p_{T} \) are conditionally independent (CI) given \( p_{RAT} \) (compare to Definition 2.1). \( \square \)

**Theorem 5.1.** The class \( \mathcal{P}_Q(V) \) is nonempty if and only if the lattice \( Q \) is distributive.

**Proof.** See Appendix A.2. \( \square \)

The normal statistical model \( \mathcal{N}_V(Q) \) defined by the requirement (5.1) of pairwise conditional independence wrt \( Q \) is then given by

\[
(5.2) \quad \mathcal{N}_V(Q) := (N(\sigma) \mid \sigma \in \mathcal{P}_Q(V))
\]

(compare to (1.8)). By Theorem 5.1, \( \mathcal{N}_V(Q) \neq \emptyset \) if and only if \( Q \) is distributive.

**Example 5.1.** Let \( V = \mathbb{R}^I \), where \( I \) is a finite index set. Every subring \( \mathcal{X} \subseteq \mathcal{Z}(I) \) determines a distributive sublattice \( Q(\mathcal{X}) \subseteq \mathcal{Q}(\mathbb{R}^I) \) as follows. For
each $K \in \mathcal{I}$ define the coordinate projection $p_K: \mathbb{R}^I \to \mathbb{R}^K$ by $p_K((x_i | i \in I)) = (x_i | i \in K)$. Since $\mathcal{I}$ is a ring, it follows that $\mathcal{Q}(\mathcal{I}) := \{(\mathbb{R}^K, p_K) | K \in \mathcal{I}\}$ is a distributive lattice of quotients of $\mathbb{R}^I$. If $\emptyset, I \in \mathcal{I}$, then $\{0\}, \mathbb{R}^I \in \mathcal{Q}(\mathcal{I})$. Thus each canonical coordinate-wise CI model $N(\mathcal{I})$ given by (1.8) is a special case of the general CI model $N(\mathcal{Q})$ given by (5.2). \qed

Conversely, by Proposition 5.1 below every distributive sublattice $\mathcal{Q} \subseteq \mathcal{Q}(\mathcal{V})$ can be represented in the form $\mathcal{Q} = \mathcal{Q}(\mathcal{I})$ for some ring of subsets $\mathcal{I}$ and every CI model $N(\mathcal{Q})$ can be represented as a canonical model $N(\mathcal{I})$.

5.3. Reduction of the CI model to canonical coordinate-wise form.

Proposition 5.1. Let $\mathcal{Q} \subseteq \mathcal{Q}(\mathcal{V})$ be a distributive lattice of quotients. Then there exists a set $I$, a ring $\mathcal{I}$ of subsets of $I$ with the property $\emptyset, I \in \mathcal{I}$, a lattice isomorphism $\mathcal{Q} \to K(\mathcal{Q})$ of $\mathcal{Q} \to \mathcal{I}$, and a basis $(e_i | i \in I)$ for $V$ such that the quotients $(\mathcal{Q}, p_\mathcal{Q}) \in \mathcal{Q}$ can be represented as follows:

\begin{align}
(5.3) & \quad \mathcal{Q} = \text{span}\{e_i | i \in K(\mathcal{Q})\}, \\
(5.4) & \quad p_\mathcal{Q}(e_i) = \begin{cases} e_i & \text{for } i \in K(\mathcal{Q}) \\ 0 & \text{for } i \in I \setminus K(\mathcal{Q}). \end{cases}
\end{align}

Proof. See Appendix A.1. \qed

We say that a basis $(e_i | i \in I)$ for $V$ satisfying the conditions in Proposition 5.1 is adapted to $\mathcal{Q}$. Thus, when $V$ is identified with $\mathbb{R}^I$ through a $\mathcal{Q}$-adapted basis $(e_i | i \in I)$, the distributive lattice $\mathcal{Q} \subseteq \mathcal{Q}(\mathcal{V})$ is identified with the ring $\mathcal{I}(\mathcal{Q}) := \{K(\mathcal{Q}) | Q \in \mathcal{Q}\}$ of subsets of $I$ and the
quotients $p_Q$, $Q \in \mathcal{Q}$, are identified with the coordinate projections $p_K^I: \mathbb{R}^I \rightarrow \mathbb{R}^K$, $K \in \mathcal{A}(Q)$ (cf. Example 5.1). Furthermore, $P(V)$ is identified with $P(I)$ through the correspondence $\sigma \rightarrow \Sigma$, where $\Sigma$ is the matrix of $\sigma$ wrt the dual basis $(e_i^* | i \in I)$ for $V^*$. The condition (5.1) is then transformed into the condition (2.1), hence $P_Q(V)$ is identified with $P_{\mathcal{A}(Q)}(I)$ and the model $N_V(Q)$ is transformed into the canonical form $N(\mathcal{A}(Q))$.

Remark 5.1. Since the identity matrix $1_I \in P_{\mathcal{A}(Q)}(I)$, the model $N_V(Q)$ is nonempty when $Q$ is distributive.

5.4. Invariant formulation of the testing problem.

Let $Q$ and $\mathcal{J}$ be two distributive sublattices of $Q(V)$ such that $\mathcal{J} \subset Q$. Then $P_Q(V) \subseteq P_{\mathcal{J}}(V)$ and one may consider the general problem of testing $N_V(Q)$ against the (possible) larger model $N_V(\mathcal{J})$ on the basis of n i.i.d. observations from $V$, i.e., testing

$$H_0: \sigma \in P_Q(V) \quad \text{vs.} \quad H: \sigma \in P_{\mathcal{J}}(V).$$

By Proposition 5.1 we may choose a $Q$-adapted basis $(e_i | i \in I)$ for $V$; clearly this basis is also adapted to $\mathcal{J}$. It follows immediately that the testing problem (5.5) is transformed into the canonical testing problem (4.1) by this choice of a $Q$-adapted basis.

§6. CONCLUDING REMARKS.

Several interesting questions remain open concerning the structure of the normal CI models $N(\mathcal{A})$ and the associated testing problems. Among these is the question under investigation is that of characterizing the
minimal determining sets of CI conditions for $\mathcal{N}(\mathfrak{X})$ (cf. Remarks 3.2 and 4.1). A second question is whether every testing problem of the general form (4.1) can be decomposed into a product of simpler testing problems (cf. Remark 4.2). 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 (4.1).

The normal statistical models $\mathcal{N}(\mathfrak{X})$ may be generalized in several ways. One natural and possibly fruitful extension is suggested by an examination of the $\mathfrak{X}$-parametrization (2.32) of $\mathcal{P}_\mathfrak{X}(I)$. A large class of "second-order" submodels of $\mathcal{N}(\mathfrak{X})$ may be obtained by replacing each $\mathcal{P}([[K]])$ in (2.32) by $\mathcal{P}_\mathfrak{X}'([[K]])$, where each $\mathfrak{X}' \equiv \mathfrak{X}'(K)$ is subring of $\mathfrak{Z}([[K]])$.

Third-order and higher-order submodels may be obtained by iterating this process. This construction yields a rich 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 $\mathcal{M}([[K] \times \langle K \rangle] \times \mathcal{P}([[K]]))$ in the $\mathfrak{X}$-parametrization (2.32) by a suitable covariance selection model requirement (cf. Dempster (1972), Wermuth (1976, 1980)), thus generalizing the multivariate graphical chain models of Lauritzen and Wermuth (1989) to "multivariate graphical lattice models".

Another interesting question is the relation of the lattice CI models $\mathcal{N}(\mathfrak{X})$ (and their extensions just described) to the CI models determined by decomposable graphs (cf. Lauritzen (1985, 1989), Lauritzen and Wermuth (1989), etc.). It appears that the class of decomposable graphical CI models neither contains nor is contained in the class of lattice CI models contained here.
APPENDIX.

In Appendices A.1 and A.2, the notation and terminology of Section 5 are followed.

A.1. The Decomposition Theorem and existence of a $\mathcal{X}$-adapted basis.

Lemma A.1. For $R \in \mathcal{Q}(V)$, the set $\mathcal{Q}(V)_R := \{Q \in \mathcal{Q} | Q \leq R\}$ is a sublattice of $\mathcal{Q}(V)$ isomorphic to the lattice $\mathcal{Q}(R)$ of quotients of $R$ through the lattice isomorphism

$$\mathcal{Q}(R) \leftrightarrow \mathcal{Q}(V)_R$$

$$(Q, p_{QR}) \leftrightarrow (Q, p_{QR} \circ p_R).$$

Proof. Straightforward. 

Lemma A.2. Let $R, T \in \mathcal{Q}(V)$ with $RVT = V$ and let $r_R: R \to p_{\mathcal{RAT}}^{-1}R(0)$ and $r_T: T \to p_{\mathcal{RAT}}^{-1}T(0)$ be surjective linear mappings. Then the linear mapping

$$(A.1) \quad \varphi: V \to (\mathcal{RAT}) \times p_{\mathcal{RAT}}^{-1}R(0) \times p_{\mathcal{RAT}}^{-1}T(0)$$

$$x \to (p_{\mathcal{RAT}}(x), r_R(p_R(x)), r_T(p_T(x)))$$

is bijective.

Proof: Suppose that $\varphi(x) = 0$. Then $p_{\mathcal{RAT}}(x) = 0$ and we obtain that $p_R(x) \in p_{\mathcal{RAT}}^{-1}R(0)$. In fact $p_R(x) = 0$ since $r_R$ is surjective. Similarly $p_T(x) = 0$, hence $x \in p_R^{-1}(0) \cap p_T^{-1}(0) = \{0\}$. The linear mapping $\varphi$ is thus injective.

Since $\dim(V) = \dim((\mathcal{RAT}) \times p_{\mathcal{RAT}}^{-1}R(0) \times p_{\mathcal{RAT}}^{-1}T(0))$, $\varphi$ is also surjective. \qed
As in Section 5.3, let \( Q \) be a distributive sublattice of \( \mathcal{Q}(V) \) such that \( \{0\}, V \in \mathcal{Q}(V) \). For \( Q \in \mathcal{Q} \), \( Q \neq \{0\} \), define

\[ \langle Q \rangle := \mathcal{V}(Q' \in \mathcal{Q} | Q' < Q) \]

and let \( J(Q) \) denote the poset of all join-irreducible elements in \( Q \), i.e.,

\[ J(Q) := \{ Q \in \mathcal{Q} | Q \neq \{0\}, \langle Q \rangle < Q \} = \{ Q \in \mathcal{Q} | Q \neq \{0\}, \forall R, T \in Q: Q = RVT \Rightarrow Q = R \text{ or } Q = T \}. \]

In the following theorem the space \( V \) is represented as a product of vector spaces indexed by \( J(Q) \) such that the space with index \( Q \in J(Q) \) has dimension \( \dim(Q) - \dim(\langle Q \rangle) \).

**Theorem A.1.** (Decomposition Theorem). For each \( Q \in J(Q) \), let \( r_Q: Q \rightarrow p_{\langle Q \rangle}^{-1}Q(0) \) be any surjective linear mapping. Then the linear mapping

\[ \varphi_V: V \rightarrow X(p_{\langle Q \rangle}^{-1}Q(0) | Q \in J(Q)) \]

\[ x \rightarrow (r_Q(p_Q(x))) | Q \in J(Q) \]

is bijective.

**Proof.** For \( R \in Q \) define \( \mathcal{Q}_R := \{ Q \in \mathcal{Q} | Q \leq R \} \), a sublattice of \( \mathcal{Q} \) (\( \mathcal{Q}_V \equiv \mathcal{Q} \)). Then

\[ R = \mathcal{V}(Q \in J(\mathcal{Q}_R)) \]

\[ J(\mathcal{Q}_R) = J(Q) \cap \mathcal{Q}_R \]

\[ J(\mathcal{Q}_RAT) = J(\mathcal{Q}_R) \cap J(\mathcal{Q}_T) \]

\[ J(\mathcal{Q}_RCT) = J(\mathcal{Q}_R) \cup J(\mathcal{Q}_T) \]
(cf. (2.4) - (2.7)). The proof proceeds by induction on $|J(Q)| = q$. If $q = 1$, then $Q = \{\emptyset, V\}$ and the result is trivial. Next, assume that the result is true whenever $q \leq k - 1$ and suppose that $q = k$. If $V \in J(Q)$ then $|J(Q_{<V}>)| = k - 1$, hence the mapping

$$<V> \rightarrow X(p_{<Q>,Q(0)}^{-1}|Q \in J(Q_{<V}>))$$

$$x \rightarrow (r_{Q}(p_{Q,<V>(x)}))|Q \in J(Q_{<V>})$$

is bijective by the induction assumption and Lemma A.1. Since the linear mapping

$$V \rightarrow <V> \times p_{<V>}^{-1}(0)$$

$$x \rightarrow (p_{<V>(x)}, r_{V}(x))$$

is bijective and $p_{Q,<V>o}p_{<V>} = p_{Q}$ for every $Q \in J(Q_{<V>})$, the mapping (A.2) is bijective in this case.

If, on the other hand, $V \notin J(Q)$, then $V = R \cup T$ where $R < V$ and $T < V$. It follows from (A.3) that $|J(Q_{R})| < k$ and $|J(Q_{T})| < k$, so by the induction assumption and Lemma A.1, the mapping

$$V \rightarrow X(p_{<Q>,Q(0)}^{-1}|Q \in J(Q_{R}))$$

$$x \rightarrow (r_{Q}(p_{Q}(x)))|Q \in J(Q_{R})$$

is (equivalent to) the quotient mapping $p_{R}: V \rightarrow R$. Similarly, the quotient mappings $p_{T}$ and $p_{RAT}$ can be represented in an analogous way, hence

$$p_{RAT,R}(0) = X(p_{<Q>,Q(0)}^{-1}|Q \in J(Q_{R}) \backslash J(Q_{RAT}))$$

$$p_{RAT,T}(0) = X(p_{<Q>,Q(0)}^{-1}|Q \in J(Q_{T}) \backslash J(Q_{RAT}))$$. 
Thus, by (A.5) and (A.6),

\[ X(p^{-1}_{Q, Q}(0) | Q \in J(Q)) = X(p^{-1}_{Q, Q}(0) | Q \in J(Q_{RAT})) \]
\[ \times X(p^{-1}_{Q, Q}(0) | Q \in J(Q_{R}) \setminus J(Q_{RAT})) \]
\[ \times X(p^{-1}_{Q, Q}(0) | Q \in J(Q_{T}) \setminus J(Q_{RAT})). \]

Lemma A.2 now implies that \( \varphi_V \) is bijective. \( \Box \)

**Remark A.1.** The representation (A.2) shows that \( V \) can be identified with a product of vector spaces indexed by \( J(Q) \); similarly, each \( R \in Q \) can be identified with the product \( X(p^{-1}_{Q, Q}(0) | Q \in J(Q_{R})) \) through the bijective linear mapping \( \varphi_R \) defined by \( \varphi_R(x) = (r_Q(p_Q(x)) | Q \in J(Q_{R})) \), \( x \in R \); under these identifications, each mapping \( p_{RT} \), \( R \leq T \leq V \), is simply a canonical projection mapping. \( \Box \)

**Proof of Proposition 5.1.** For each \( Q \in J(Q) \), let \( [K(Q)] \) be a set with

\[ |[K(Q)]| = \dim(p^{-1}_{Q, Q}(0)). \]

For \( R \in Q \), define

\[ (A.7) \quad K(R) := \bigcup ([K(Q)] | Q \in J(Q_{R})) \]

and define \( I := K(V) \). From (A.5) and (A.6) it follows that \( \mathcal{A} = \mathcal{A}(Q) := \{ K(R) | R \in Q \} \) is a subring of \( \mathcal{A}(I) \) and the mapping \( R \to K(R) \) is a lattice isomorphism between \( Q \) and \( \mathcal{A} \). Now Remark A.1 implies that there exists a basis \( \{ e_i \mid i \in I \} \) for \( V \) such that the elements \( (R, p_R) \) in \( Q \) can be represented as in (5.3) and (5.4). \( \Box \)
A.2. Proof of Theorem 5.1.

Lemma A.3. Suppose that \( x \sim N(\sigma), \sigma \in \mathbb{P}(V) \). Then for any \( R, T \in \mathbb{Q}(V) \),

\[
p_R(x) \parallel p_T(x) |p_{\text{RAT}}(x) \iff p_R^{-1}(0) \text{ and } p_T^{-1}(0) \text{ are geometrically orthogonal (g.o.) wrt the inner product } \delta := \sigma^{-1} \text{ on } V \text{ (cf. [A] (1990), Definition 4.1, for the definition of g.o.).}
\]

Proof. Let \( p_R^{-1}(0)^\perp, p_T^{-1}(0)^\perp, \) and \( p_{\text{RAT}}^{-1}(0)^\perp \) denote the orthogonal complements of \( p_R^{-1}(0), p_T^{-1}(0), \) and \( p_{\text{RAT}}^{-1}(0), \) respectively, wrt \( \delta \).

Furthermore, let \( q_R, q_T, \) and \( q_{\text{RAT}} \) be the orthogonal projections of \( V \) onto \( p_R^{-1}(0)^\perp, p_T^{-1}(0)^\perp, \) and \( p_{\text{RAT}}^{-1}(0)^\perp \). Then \( (p_R^{-1}(0)^\perp, q_R), (p_T^{-1}(0)^\perp, q_T) \) and \( (p_{\text{RAT}}^{-1}(0)^\perp, q_{\text{RAT}}) \) represent the quotients \((R, p_R), (T, p_T), \) and \((\text{RAT}, p_{\text{RAT}}), \) respectively. Therefore

\[
p_R(x) \parallel p_T(x) |p_{\text{RAT}}(x)
\]

\[
\iff q_R(x) \parallel q_T(x) |q_{\text{RAT}}(x)
\]

\[
\iff (q_R(x) - q_{\text{RAT}}(x)) \parallel (q_T(x) - q_{\text{RAT}}(x)) |q_{\text{RAT}}(x)
\]

\[
\iff (q_R(x) - q_{\text{RAT}}(x)) \parallel (q_T(x) - q_{\text{RAT}}(x)).
\]

\[
\iff (p_R^{-1}(0)^\perp \cap p_{\text{RAT}}^{-1}(0)) \perp (p_T^{-1}(0)^\perp \cap p_{\text{RAT}}^{-1}(0))
\]

\[
\iff p_R^{-1}(0)^\perp \text{ and } p_T^{-1}(0)^\perp \text{ are g.o.}
\]

\[
\iff p_R^{-1}(0) \text{ and } p_T^{-1}(0) \text{ are g.o.}
\]

The third \( \iff \) follows since \((*) \) \( (q_R - q_{\text{RAT}}, q_T - q_{\text{RAT}}) \) is a projection onto \( (p_R^{-1}(0)^\perp \cap p_{\text{RAT}}^{-1}(0)) \otimes (p_T^{-1}(0)^\perp \cap p_{\text{RAT}}^{-1}(0)) \), and this direct sum is orthogonal to \( p_{\text{RAT}}^{-1}(0)^\perp \) wrt \( \delta \). The fourth \( \iff \) follows from \((*)\), while the fifth and sixth \( \iff \)'s are elementary properties of geometric orthogonality. \( \square \)
Proof of Theorem 5.1. Since the correspondence \( Q \leftrightarrow P_Q^{-1}(0) \) between \( \mathcal{L}(V) \) and the lattice \( \mathcal{L}(V) \) of all subspaces of \( V \) (cf. [A] (1990), Section 4.1) is a lattice anti-isomorphism it follows that \( \mathcal{L} := \{ P_Q^{-1}(0) \mid Q \in \mathcal{L} \} \subseteq \mathcal{L}(V) \) is a lattice and is anti-isomorphic to \( \mathcal{Q} \). If \( \sigma \in P_Q(V) \neq \emptyset \), then by Lemma A.3, \( \mathcal{L} \) is g.o. wrt \( \delta := \sigma^{-1} \). Thus by Proposition 4.1 of [A] (1990) \( \mathcal{L} \) is distributive, hence so is \( \mathcal{Q} \). Conversely, if \( \mathcal{Q} \) is distributive, then \( P_Q(V) \neq \emptyset \) by Remark 5.1.

A.3. Proof of Theorem 4.2.

Let \( \Omega \subseteq M(I \times N) \) be the open subset

\[
\Omega := \{ y \in M(I \times N) \mid \text{rank}(y) = \min(|I|, n) \}.
\]

Since \( M(I \times N) \setminus \Omega \) is a Lebesgue-null set, we may replace the sample space \( M(I \times N) \) by \( \Omega \). Also, since \( \text{rank}(Ay) = \text{rank}(y) \) for \( A \in GL_n(I) \) and \( y \in M(I \times N) \), it follows that \( GL_n(I) \) acts on \( \Omega \) by restriction of (3.12).

Furthermore, since \( \Omega \) is locally compact, Lemma A.5 at the end of this subsection implies that this restriction is a proper action (whereas (3.12) itself is not proper). Thus, in order to prove Theorem 4.2 we may apply the method of [A] (1982) to study the transformation of the normal distributions in the model \( H_0 \) under the mapping

\[
\begin{align*}
\Omega & \rightarrow \Omega/\text{GL}_n(I) \times (X(P([K]) \mid K \in J(\mathfrak{A}))) \\
y & \mapsto (\pi(y), \{ \mathcal{L}_{[K]}(y) \mid K \in J(\mathfrak{A}) \}).
\end{align*}
\]

The group \( GL_n(I) \) is the semidirect product of its two closed subgroups \( \mathcal{A} \) and \( J \), where
\[ \mathcal{d} := \{ A \in \text{GL}_d(I) \mid A_{[K]} = 0, \ K \in \mathcal{J}(d) \} \]

\[ \mathcal{J} := \{ T \in \text{GL}_d(I) \mid T_{[K]} = 1_{[K]}, \ K \in \mathcal{J}(d) \}. \]

Therefore we may apply the method of \cite{A} (1982), Section 5, with \( K = \text{GL}_d(I), \ H = \mathcal{d}, \ G = \mathcal{J}, \) and \( X = \Omega \) to see that \( \tau \) can be represented as \( \tau = \tau_{\mathcal{d}} \circ \tau_{\mathcal{J}}, \) where \( \tau_{\mathcal{J}} : \Omega \rightarrow \Omega/\mathcal{J} \) and \( \tau_{\mathcal{d}} : \Omega/\mathcal{J} \rightarrow (\Omega/\mathcal{J})/\mathcal{d} \simeq \Omega/\text{GL}_d(I). \) (The action of \( \mathcal{J} \) on \( \Omega \) is the restriction of (3.12) to \( \mathcal{J} \times \Omega, \) and the induced action of \( \mathcal{d} \) on \( \Omega/\mathcal{J} \) is defined as in equation (21) of \cite{A} (1982).)

Since the mapping (A.9) is invariant under the action of \( \mathcal{J} \) on \( \Omega \) (cf. (2.26)), it has a unique factorization through \( \tau_{\mathcal{J}}. \) Therefore we may first transform the normal distributions in the model \( H_0 \) from \( \Omega \) to \( \Omega/\mathcal{J} \) by \( \tau_{\mathcal{J}}. \)

To do this, we need the following explicit representation:

**Lemma A.4.** A representation of \( \tau_{\mathcal{J}} : \Omega \rightarrow \Omega/\mathcal{J} \) is given by

\[ (A.10) \quad \Omega/\mathcal{J} = \{ y \in \Omega \mid y_{[K]} y_{[K]}^t (y_{[K]} y_{[K]}^t)^{-1} y_{[K]} = 0, \ K \in \mathcal{J}(d) \}. \]

\[ (A.11) \quad \tau_{\mathcal{J}}(y) = (y_{[K]} - y_{[K]} y_{[K]}^t (y_{[K]} y_{[K]}^t)^{-1} y_{[K]} \mid K \in \mathcal{J}(d)). \]

**Proof:** To show that \( \Omega/\mathcal{J} \) in (A.10) is a cross-section of \( \Omega \) and that \( \tau_{\mathcal{J}} \) in (A.11) is a maximal invariant function, it suffices to show that for each \( y \in \Omega, \)

\[ (A.12) \quad \{ Ty \mid T \in \mathcal{J} \} \cap (\Omega/\mathcal{J}) = \{ \tau_{\mathcal{J}}(y) \}. \]

To show the inclusion \( \subseteq \) in (A.12), suppose that \( Ty \in \Omega/\mathcal{J}. \) Then from (A.10), Proposition 2.2(ii), and (2.18),
for each $K \in J(\mathcal{X})$, hence

\begin{equation}
(A.13)
(T_y)[K] = y[K] - y[K]y^t <K>; (y^t <K>)^{-1} y <K>.
\end{equation}

$K \in J(\mathcal{X})$, i.e., $-T_y = \pi_y(y)$. To show the opposite inclusion \(\exists\), it is easy to verify that $\pi_y(y) \in (\Omega/\mathcal{Y})$ for every $y \in \Omega$; to show that $\pi_y(y) \in 
\{Ty|T \in \mathcal{Y}\}$, simply note that $\pi_y(y) = \tilde{T}_y$ where

\[\tilde{T}_y[K] = -y[K]y^t <K>; (y^t <K>)^{-1}.\]

$K \in J(\mathcal{X})$. Finally, the mapping $\pi_y : \Omega \rightarrow \Omega/\mathcal{Y}$ defined in (A.10) and (A.11) is clearly continuous, so this representation is also topological and the result follows. \(\square\)

We may now apply formula (16) of [A] (1982) to transform the normal distributions in the model $H_0$ by the mapping $\pi_y$ given by (A.11). In the notation of Section 4 of [A] (1982), $G = \mathcal{Y}$, $X = \Omega$, $\lambda$ is the restriction of Lebesgue measure on $\mathcal{M}(1\times N)$ to the open subset $\Omega$, $\Pi = \pi_y$, $\beta$ is a Haar measure on $\mathcal{Y}$, $\Lambda_G = \Lambda_y \equiv 1$, and $P = p\cdot\lambda$, i.e., $P$ is the normal distribution with density $p$ given by

\[p(y) = (\det(\Sigma))^{-n/2}\exp\{-tr(\Sigma^{-1}yy^t)/2\}, \quad y \in \Omega,\]
with respect to \( \lambda \). For \( \Sigma \in P_{\mathcal{A}}(\mathcal{I}) \), the density \( q \) of \( \pi_{\mathcal{I}}(\mathcal{P}) \) wrt the quotient measure \( \lambda/\beta \) on \( \Omega/\mathcal{I} \) is thus given by

\[
(A.14) \quad q(\pi_{\mathcal{I}}(y)) = (\det(\Sigma))^{-n/2} \int_{\mathcal{I}} \exp\{-\text{tr}(\Sigma^{-1}(Ty)(Ty)^t)/2\} d\beta(T)
\]

\[
= II((\det(\Sigma_{[K]}))^{-n/2}\exp\{-\text{tr}(\Sigma_{[K]}^{-1}y_{[K]}^t)y_{[K]}^t)/2\} | K \in J(\mathcal{I}))
\]

\[
\times II(\exp\{-\text{tr}(\Sigma_{[K]}^{-1}z_{[K]}(T)z_{[K]}(T)^t)/2\} | K \in J(\mathcal{I}))d\beta(T),
\]

where

\[
y_{[K]}^t = (y_{[K]} - y_{[K]}^t y_{[K]} y_{[K]}^t)^{-1} y_{[K]}.
\]

\[
z_{[K]}(T) = (y_{[K]} + T_{\mathcal{K}}y_{\mathcal{K}} y_{\mathcal{K}}^t y_{\mathcal{K}} y_{\mathcal{K}}^{-1} y_{\mathcal{K}} - \Sigma_{[K]} y_{\mathcal{K}} y_{\mathcal{K}}^t y_{\mathcal{K}}^t)^{-1} y_{\mathcal{K}}^t T_{\mathcal{K}}y_{\mathcal{K}} y_{\mathcal{K}}^t.
\]

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

Since \( d\beta(T) = II(d\lambda_K(T_{\mathcal{K}}) | K \in J(\mathcal{I})) \), where \( \lambda_K \) is the Lebesgue measure on \( M([K] \times \mathcal{K}) \) (cf. (2.19), the last integral in (A.14) can be calculated using Fubini's Theorem and the translation invariance of \( \lambda_K, K \in J(\mathcal{I}). \)

The order of integration should be determined by a never-increasing listing \( K_1, K_2, \ldots, K_{|J(\mathcal{I})|} \) of the elements in \( J(\mathcal{I}) \) (cf. Remark 2.1). After some calculation we obtain

\[
(A.15) \quad q(\pi_{\mathcal{I}}(y))
\]

\[
= II((\det(\Sigma_{[K]}))^{-n/2}\exp\{-\text{tr}(\Sigma_{[K]}^{-1}y_{[K]}^t)y_{[K]}^t)/2\} | K \in J(\mathcal{I}))
\]

\[
\times II((\det(\Sigma_{[K]}))^{\mathcal{K}}/2(\det(y_{\mathcal{K}} y_{\mathcal{K}}^t))^{-1}|K|/2 | K \in J(\mathcal{I}))
\]
\[ = \Pi((\det(\Sigma_{[K]}))^{-(n-|K|)/2} \exp(-\text{tr}(\Sigma^{-1} S_{[K]} y)/2) |K \in J(x)) \]

\[ \times \Pi((\det S_{[K]}(y))^{-|K|/2} |K \in J(x)), \quad \pi_f(y) \in \Omega/\mathcal{F}, \]

where \( S(y) = y^t. \)

By Lemma A.4, where \( \Omega/\mathcal{F} \) is represented as a subset of \( \Omega \), the induced action of the subgroup \( \mathcal{A} \) on \( \Omega/\mathcal{F} \) is simply the restriction of the action (3.12) to \( \mathcal{A} \times (\Omega/\mathcal{F}) \). The next step is to represent the transformed measure \( \pi_f(P) = q^*(\lambda/\beta) \) as \( \pi_f(P) = q_1 \cdot \nu \), where \( \nu \) is an invariant measure under this action of \( \mathcal{A} \) on \( \Omega/\mathcal{F} \).

It follows from the statement following the proof of Proposition 2 on p. 961 of [A] (1982) that the quotient measure \( \lambda/\beta \) is relatively invariant under the action of \( \mathcal{A} \) on \( \Omega/\mathcal{F} \) with multiplier \( \chi \) given by \( \chi(A) = (\mod \varphi_A)^{-1} \chi_0(A), \ A \in \mathcal{A} \), where \( \chi_0 \) is the multiplier for \( \lambda \) as a relatively invariant measure under the action of \( \text{GL}_x(I) \) on \( \Omega \) and where the automorphisms \( \varphi_A: \mathcal{F} \to \mathcal{F} \) are defined by \( \varphi_A(T) = A T A^{-1}, \ T \in \mathcal{F} \). Since \( A = \text{Diag}(A_{[K]} | K \in J(x)) \) it is clear that

\[ (\varphi_A(T))_{[K]} = A_{[K]} T_{[K]} A_{[K]}^{-1} \]

\[ \forall \ K \in J(x). \]

hence

\[ \mod \varphi_A = \Pi(\left|\det(A_{[K]})\right|^{n-|K|} \left|\det(A_{[K]}^{-1})\right| \left|\det(A_{[K]}^{-1})\right| |K \in J(x)). \]

However,

\[ \chi_0(A) = |\det(A)|^n = \Pi(\left|\det(A_{[K]})\right|^{n} |K \in J(x)), \]

so that

\[ \chi(A) = \Pi(\left|\det(A_{[K]})\right|^{n-|K|} \left|\det(A_{[K]}^{-1})\right| \left|\det(A_{[K]}^{-1})\right| |K \in J(x)). \]
If we define $m: \mathbb{N} \to ]0, \infty[$ by

$$m(y) = \Pi((\det(A_{[K]}(y)))^{(n-|K|)/2} \cdot \det(S_{<K>(y)}) |K|/2 |K \in J(\mathfrak{A})),$$

it follows that $m(Az) = \chi(A) m(z)$, $z \in \mathbb{N}$, $A \in \mathfrak{A}$ (compare to (17) in [A] 1982). Thus the measure $\nu := m^{-1}(\lambda/\beta)$ is invariant under the action of $\mathfrak{A}$ on $\mathbb{N}$. From (A.15), the density $q_1 := mq$ of $\pi_{\mathfrak{A}}(P)$ with respect to $\nu$ is therefore given by

$$q_1(\pi_{\mathfrak{A}}(x)) =$$

$$\Pi\left(\frac{\det(A_{[K]}(y))}{\det(S_{[K]}(y))}\right)^{(n-|K|)/2} \cdot \exp\{-ntr(S^{-1}_{[K]}A_{[K]}(y)/2) |K \in J(\mathfrak{A})\},$$

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

The final step in the proof of Theorem 4.2 is to obtain the transformation of the measure $\pi_{\mathfrak{A}}(P) \equiv q_1 \cdot \nu$ under the mapping

(A.16) \[ \mathbb{N} \to (\mathbb{N})/\mathfrak{A} \times (X(P([K]) |K \in J(\mathfrak{A}))) \]

$$\pi_{\mathfrak{A}}(y) \to (\pi_{\mathfrak{A}}(\pi_{\mathfrak{A}}(y)), (A_{[K]}(y) |K \in J(\mathfrak{A}))).$$

Since the action of $\mathfrak{A}$ on $\mathbb{N}$ is the restriction to the closed subset $\mathfrak{A} \times (\mathbb{N})$ of the proper action of $GL_{\mathfrak{A}}(I)$ on $\mathbb{N}$, 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 $\kappa$ on $(\mathbb{N})/\mathfrak{A}$ such that the invariant measure $\nu$ is transformed into the product measure $\kappa \otimes \nu_0$ under the mapping (A.16), where $\nu_0$ is an invariant measure on $X(P([K]) |K \in J(\mathfrak{A}))$ under the proper and transitive action.
(A.17) \[ \delta \times (X(P([K]) | K \in J(\mathfrak{A})) \rightarrow X(P([K]) | K \in J(\mathfrak{A})) \]

\[ (A. \ (A_{[K]} | K \in J(\mathfrak{A})) \rightarrow (A_{[K]} A_{[K]}^t | K \in J(\mathfrak{A})). \]

(Lemma 3 of Andersson, Brøns, and Jensen (1983) is applied with \( G = \mathfrak{A}, X = \Omega/\mathcal{I}, Y = X(P([K]) | K \in J(\mathfrak{A})), t = (\pi_{\mathcal{I}}(y) \rightarrow (A_{[K]}^\mathcal{I}(y) | K \in J(\mathfrak{A}))), \pi = \pi_{\mathcal{I}}, \text{ and } v = v. \)

Since \( q_1(z) \) depends on \( z := \pi_{\mathcal{I}}(y) \) only through \( (A_{[K]}^\mathcal{I}(y) | K \in J(\mathfrak{A})) \), the probability measure \( q_1 \cdot v \) is therefore transformed under (A.16) into the probability measure \( r \cdot (\kappa \Phi_v) \), where

\[ r : (\Omega/\mathcal{I})/\mathfrak{A} \times (X(P([K]) | K \in J(\mathfrak{A})) \rightarrow \mathbb{R}_+ \]

\[ (w, (A_{[K]} | K \in J(\mathfrak{A})) \rightarrow \]

\[ \Pi \left[ \frac{\det(A_{[K]})}{\det(A_{[K]}^\mathcal{I})} \right]^{(n-|K|)/2} \times \exp \left\{ -n \text{tr} \left( A_{[K]}^\mathcal{I} A_{[K]} / 2 \right) | K \in J(\mathfrak{A}). \right\} \]

Because \( r \) does not depend on \( w \) under \( H_0 \) it follows that \( \pi = \pi_{\mathcal{I}} \circ \pi_{\mathcal{J}} \) is independent of \( (A_{[K]}^\mathcal{I} | K \in J(\mathfrak{A})) \). \( \pi \) has distribution \( \kappa \), and \( (A_{[K]}^\mathcal{I} | K \in J(\mathfrak{A})) \) has distribution \( s \cdot v_0 \), where \( s((A_{[K]} | K \in J(\mathfrak{A})) \) is given by the product (A.18). Furthermore, since \( v_0 = \Theta(v_K | K \in J(\mathfrak{A})) \) where \( v_K \) is an invariant measure on \( P([K]) \) under the usual action of \( \text{GL}([K]) \), it follows that under \( H_0 \), \( A_{[K]}^\mathcal{I}, K \in J(\mathfrak{A}), \) are mutually independent and \( A_{[K]}^\mathcal{I} \) has the Wishart distribution on \( P([K]) \) with \( n-|K| \) degrees of freedom and expected value \( A_{[K]}^\mathcal{I} \). This ends the proof of Theorem 4.2.

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 A.5. 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 $\varphi(gx) = \varphi(g)\psi(x)$, $x \in X$, $g \in G$. If $\varphi$ 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

\[
\begin{array}{ccc}
G \times X & \xrightarrow{\varphi} & X \times X \\
\varphi \times \psi \downarrow & & \downarrow \psi \times \psi \\
G' \times X' & \xrightarrow{\psi'} & X' \times X'.
\end{array}
\]

where $\theta(g, x) = (gx, x)$ and $\theta'(g', x') = (g'x', x')$. We must show that $\theta^{-1}(C)$ 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., $\theta' (\varphi \times \psi) = (\psi \times \psi) \circ \theta$, it follows that

\[
\theta^{-1}(C) \subseteq \theta^{-1}( (\varphi \times \psi)^{-1}( (\psi \times \psi)(C)) ) = (\varphi \times \psi)^{-1}(\theta'^{-1}( (\psi \times \psi)(C)) ) \\
\subseteq (\varphi \times \psi)^{-1}( p_G \cdot (\theta'^{-1}( (\psi \times \psi)(C)) \times X ) ) = \varphi^{-1}(C') \times X
\]

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

\[
\theta^{-1}(C) \subseteq \varphi^{-1}(C') \times p_2(C).
\]

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

With the identifications $G = G' = \text{GL}_A(I)$, $X = \Omega$, $X' = \text{P}_A(I)$, $\varphi$ is the identity mapping on $\text{GL}_A(I)$, and $\psi = \frac{A}{2}$. Lemma A.5 may be applied as indicated at the beginning of this subsection.
REFERENCES


likelihood in mixed graphical interaction models. *Biometrika* 76,
539-555.

J. Statist.* 7, 53-60.


Lauritzen, S.L. (1985). Test of hypotheses in decomposable mixed

Statist.* 16, 273-306.

Independence Properties of Directed Markov Fields. To appear in
*Networks*.

Lauritzen, S.L. and Wermuth, N. (1989). Graphical models for association between variables, some of which are qualitative and some quantitative. 


