S. A. Andersson M. D. Perlman

Lattice-Ordered Conditional Independence Models for Missing Data

Preprint
October
1989
9

Institute of Mathematical Statistics
University of Copenhagen

ISSN 0902-8846

Steen A. Andersson and Michael D. Perlman

LATTICE-ORDERED CONDITIONAL INDEPENDENCE MODELS FOR MISSING DATA

Preprint 1989 No. 9

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

October 1989

LATTICE-ORDERED CONDITIONAL INDEPENDENCE MODELS FOR MISSING DATA[1]

Steen A. Andersson[2] and Michael D. Perlman[3]

ABSTRACT

Statistical inference for the parameters of a multivariate normal distribution $N_p(\mu,\Sigma)$ based on a sample with missing observations is straightforward when the missing data pattern is monotone (= nested), reducing to the analysis of several normal linear regression models by step-wise conditioning. When the missing data pattern is non-monotone, however, such analysis is impossible. It is shown here that *every* missing data pattern naturally determines a set of lattice-ordered conditional independence restrictions which, when imposed upon the unknown covariance matrix Σ, yields a factorization of the joint likelihood function as a product of (conditional) likelihood functions of normal linear regression models just as in the monotone case. From this factorization the maximum likelihood estimators of μ and Σ (under the conditional independence restrictions) can be explicitly derived.

---

[1] This research was supported in part by the U.S. National Science Foundation Grant No. DMS 89-02211.

[2] Institute of Mathematical Statistics, University of Copenhagen, Denmark, and Department of Mathematics, University of Indiana, Bloomington, Indiana 47401.

[3] Department of Statistics, University of Washington, Seattle, Washington 98195.

1. Introduction.

Suppose that $x_1, \cdots, x_n$ represent a sample of stochastically independent random vectors from a p-variate normal distribution $N(\mu,\Sigma)$ with mean vector μ and positive definite covariance matrix Σ, both unknown. Each x_j and μ are p-dimensional column vectors. Frequently in practice, some of the p components of one or more x_j are unobserved or missing. Thus the observed data array may assume forms such as the following four examples, where in each case $p = 2$, $n = 5$:

$$\begin{array}{ccccc}1&1&1&1&1\\2&2&2&2&2\end{array}\qquad\qquad \begin{array}{ccccc}1&1&1&1&1\\&2&&&2\end{array}\qquad\qquad \begin{array}{ccccc}&1&1&1&\\2&2&2&&2\end{array}\qquad\qquad \begin{array}{ccccc}1&&&1&1\\&2&2&&\end{array}$$

Figure 1.1.

In each array a "1" ("2") indicates that the first (second) component of that column vector x_j is present, while a blank indicates a missing observation[1].

After permuting columns and combining identical columns, it is seen that the data arrays in Figure 1.1 determine the following four incomplete data patterns:

$$\begin{array}{c}1\\2\end{array}\qquad\qquad \begin{array}{cc}1&1\\&2\end{array}\qquad\qquad \begin{array}{ccc}1&&1\\&2&2\end{array}\qquad\qquad \begin{array}{cc}1&\\&2\end{array}$$

monotone (complete)	monotone (Σ identifiable)	non-monotone (Σ identifiable)	non-monotone (σ_{12} unidentifiable)

Figure 1.2.

Each pattern is specified by the class $\mathbf{S}$ of subsets of indices determined by its columns, so the four patterns in Figure 1.2 are respectively equivalent to the four classes

(1.1) $$\{12\},\quad \{1, 12\},\quad \{1, 2, 12\},\quad \{1, 2\},$$

---

[1]More generally, the entries "1", "2", $\cdots$, "q", in such arrays may represent multivariate columns of variates with every column labelled "i" having the same dimension p_i, where $p_1 + \cdots + p_q = p$.

where "12" denotes the subset {1, 2}, etc.

1.1. Monotone and non-monotone incomplete data patterns.

A incomplete data pattern is called *monotone* (= nested, hierarchical, stair-case, etc.) if the p variates can be relabelled such that if variate i is missing in vector x_j, then the variates $i+1, \cdots, p$ are also missing in x_j. Equivalently, $\mathcal{S}$ is monotone if its members are totally ordered by inclusion. The first two patterns in Figure 1.2 are monotone, while the last two are non-monotone. The correlation between variates 1 and 2 is unidentifiable (hence inestimable) in the fourth pattern since these two variates are never observed simultaneously. Up to permutation of rows and columns (i.e., relabelling of variates and samples) the four patterns in Figure 1.2 are the only possible incomplete data patterns for bivariate data (p = 2).

For trivariate data (p = 3), however, there are 32 possible incomplete data patterns, of which 4 are monotone and 28 are non-monotone. Some examples are given in Figure 1.3 and 1.4a, b:

```
1      11      11      111
2       2      22       22
3       3       3        3
```

Figure 1.3: The four monotone incomplete data patterns when p = 3.

```
11 1      111      1 1      1 11      1 1
2 22      2 2      2 2        22       22
 333       33       33       3 3        3
```

Figure 1.4a: Five non-monotone incomplete data patterns when p = 3; complete observations present, Σ identifiable.

1 1	1 1	1	1 1
2 2	2	2	2
33	3	3	3

Figure 1.4b: Four non-monotone incomplete data patterns when p = 3; no complete observations present.

Note that Σ is identifiable in the first pattern in Figure 1.4b even though no complete observations are present, since every pair of the variates 1, 2, 3 are observed together, whereas Σ is not identifiable in the last three patterns of Figure 1.4b.

1.2. Statistical inference for missing data models.

It is well known that statistical inference for *monotone* missing data models is relatively simple (cf. Anderson (1957), Bhargava (1962, 1975), Little and Rubin (1987), Rao (1956), and many others listed in Kariya, Krishnaiah, and Rao (1983)). Not only is (μ,Σ) identifiable since complete observations are present, but more importantly, the joint likelihood function can be factored as a product of conditional likelihood functions each having the form of an ordinary multivariate normal linear regression model. This is accomplished by factoring the joint density function f of the observed data array in the form

$$f = f(1)f(2 \mid 1)\cdots f(p \mid 1\cdots(p-1)), \tag{1.2}$$

where $f(i \mid 1\cdots(i-1))$ denotes the conditional density of all observations on variate i given the values of all observations on variates $1,\cdots,i-1$ (also recall Footnote 1). Furthermore, the factor $f(i \mid 1\cdots(i-1))$ depends on (μ,Σ) only through the usual regression parameters that appear in the conditional distribution of variate i given variates $1,\cdots,i-1$, and the full parameter space of (μ,Σ) is in 1-1 correspondence with the product of the parameter spaces of these p sets of regression parameters. (For $i = 1$, the regression parameters are simply μ_1 and Σ_{11}, the unconditional mean and variance (or covariance matrix) of the first variate (or first block of variates).) Rubin (1974) refers to these as p sets of "distinct" parameters.

For a monotone incomplete data pattern these factorizations of the likelihood function and parameter space reduce the problem of maximum likelihood estimation of (μ,Σ) to that of estimating the parameters of several linear regression models. In particular, this provides simple necessary and sufficient conditions in terms of the sample size for existence and uniqueness of maximum likelihood estimates (MLE), and also provides explicit expressions for these MLE.

For general dimension $p \geq 3$, however, the vast majority of incomplete data patterns are *non-monotone*, in which case the likelihood function and the parameter space cannot be factored simply as in the monotone case and the estimation problem cannot be reduced to a set of linear regression problems (cf. Rubin (1974) and Rubin (1987, §5.6)). The parameter Σ may or may not be identifiable, conditions for existence and uniqueness of the MLE of (μ,Σ) are not expressible in convenient form[2], and explicit expressions for the MLE are not available. In practice it is usual to apply the EM algorithm or other algorithms to approximate the MLE[3] (cf. Little and Rubin (1987, Chapter 8), Rubin (1987, §5.6)), but the EM algorithm may not converge to an unique solution, if at all, and the resulting estimates may depend heavily upon the initial value chosen for (μ,Σ) (Murray (1977)). Other proposed approximate methods may not yield positive definite estimates of Σ (Hocking and Smith (1968)). Only one proposed method, that of discarding some observations to obtain a monotone incomplete data pattern (Rubin (1974) and Rubin (1987, pp.189-190)) yields explicit MLE for (μ,Σ), but this incurs a loss of efficiency that may be substantial unless most observations are complete.

1.3. Pairwise conditional independence models for incomplete data.

In this paper we present an alternative approach to the analysis of non-monotone incomplete data patterns in a sample from a multivariate normal distribution. We shall show that every incomplete data pattern generates a finite distributive lattice which in turn determines a mathematically natural set of pairwise conditional independence (CI) conditions. When imposed upon Σ to produce a restricted parameter space, these CI conditions yield a statistical model that inherits the desirable properties of the monotone case described above. In particular, both the likelihood

---

[2]Even in the bivariate case the likelihood function may have multiple maxima - cf. Murray (1977).

[3]Sometimes without requiring that Σ be identifiable.

function and the parameter space can be factored so that the MLE of (μ,Σ) under the CI condition are obtained by solving a set of ordinary multiple linear regression models, exactly as in the monotone case. This immediately provides simple conditions for existence and uniqueness of the MLE of (μ,Σ) under the CI model and explicit expressions for the MLE when it exists.

Rubin (1987, p. 190) explicitly suggested this approach for the analysis of a simple non-monotone incomplete data pattern[4], namely:

```
11111
 2222
  3 3
   44
```

Figure 1.5.

Essentially, he notes that if it is assumed that variates 3 and 4 are CI given variates 1 and 2, which we express as $3 \perp\!\!\!\perp 4 \mid (1,2)$ following Dawid (1980), then the joint density f of the observed data array may be factored as

$$f = f(1)\,f(2 \mid 1)\,f(3 \mid 12)\,f(4 \mid 12) \tag{1.3}$$

with each factor corresponding to a standard linear regression model, whereas without the CI assumption no such factorization is possible.

Earlier, Anderson (1957) considered the following two examples of non-monotone incomplete data patterns:

```
11       111
2         22
 3       3
          4
           5
```

Figure 1.6.

[4]Our labelling of variates is different than, but equivalent to, Rubin's.

For the first pattern Anderson (1957) and, previously, Lord (1955) noted that the joint likelihood function f may be factored by straightforward sequential conditioning as follows:

$$f = f(1)f(2 \mid 1)f(3 \mid 1) \tag{1.4}$$

but they did not relate this factorization to a CI assumption. Clearly, however, the factorization (1.4) is equivalent to the CI condition $2 \perp\!\!\!\perp 3 \mid 1$.

Although giving no explicit description of his factorization procedure, Anderson (1957) states that "other problems of missing observations (but not all) can be handled in this way", including the second pattern in Figure 1.6, for which he did not state the factorization but which is easily found by "sequential conditioning" to be

$$f = f(1)f(2 \mid 1)f(3 \mid 1)f(4 \mid 12)f(5 \mid 12) \tag{1.5}$$

where again each factor is the likelihood function of a standard linear regression model. If, however, the patterns in Figure 1.6 are replaced by the augumented patterns

```
111       1111
2 2        222
 33       3  3
           4
            55
```

Figure 1.7.

then no factorization of the likelihood function can be obtained by Anderson's sequential conditioning approach, even allowing relabelling of the five variates. Indeed, without the imposition of CI conditions, no such factorizations are possible, but the appropriate CI conditions may not be readily apparent.

Application of the theory presented in Section 3 (cf. Examples 4.3 and 4.13) leads to the following minimal sets of CI restrictions that allow factorizations of the likelihood function for the two incomplete data patterns in Figure 1.7:

(1.6) First pattern: $2 \perp\!\!\!\perp 3 \mid 1$

(1.7) Second pattern: $2 \perp\!\!\!\perp 3 \mid 1$ and $3 \perp\!\!\!\perp 4 \perp\!\!\!\perp 5 \mid (1,2)$.

When these CI conditions are imposed on the covariance matrix Σ, then the factorizations (1.4) and (1.5) regain their validity for the augmented patterns.[5] In the present paper it will be shown how the lattice structure of a general incomplete data pattern $\mathfrak{S}$ determines the minimal set of CI conditions that yields factorizations such as (1.3), (1.4), and (1.5).

1.4. Applicability and limitations of CI models.

Rubin (1987, p.191) states that "in some cases, such assumptions of conditional independence may be perfectly reasonable" due to the nature of the statistical experiment. Furthermore, Andersson and Perlman (1988) show that the CI assumptions may be tested (based upon the complete observations) by standard multivariate techniques. Even in cases where the CI model is not deemed appropriate, Rubin (personal communication) has noted that the MLE of (μ,Σ) obtained under the CI model may provide useful starting values for the EM algorithm or other iterative methods for approximating the MLE under the unrestricted model. Additionally, the explicit MLE solution obtained under the CI model enables one to apply standard diagnostic methods to investigate the validity of the model assumptions.

---

[5]Thus the incomplete data patterns in Figure 1.7 lead to the same CI covariance models as the corresponding patterns in Figure 1.6, but the factorizations for Figure 1.6 may be obtained easily by inspection whereas this is not so in Figure 1.7. In the lattice-theoretic language of our general method (cf. Sections 3 and 4), this is explained by the facts that the incomplete data patterns $\mathfrak{S} = \{12, 13\}$ and $\mathfrak{S} = \{13, 124, 125\}$ in Figure 1.6 consist solely of join-irreducible elements of the corresponding lattices $\mathfrak{K} = \mathfrak{K}(\mathfrak{S})$ generated by $\mathfrak{S}$ and that both sets $J(\mathfrak{K})$ of join-irreducible elements are closed under intersection. For each of the augmented incomplete data patterns in Figure 1.7, however, the last column is not a member of $J(\mathfrak{K})$. See Examples 4.3 and 4.13 for further discussion of these patterns.

It must be noted that the lattice-ordered CI conditions imposed by a given incomplete data pattern may be severely restrictive. For example, the CI conditions determined by the first missing data pattern in Figure 1.4a require that the variates 1,2,3 must be mutually independent (cf. Example 4.7). In such cases, examination of the lattice $\mathcal{K}$ determined by the missing data pattern can show which partial observations would need to be discarded in order to obtain a less restrictive CI model (e.g., to obtain a monotone pattern, which requires no CI restrictions for explicit analysis, as Rubin (1987, pp.189-190) has suggested). Of course, efficiency considerations would be necessary to implement such a procedure.

1.5. Outline.

The lattice-ordered CI covariance models for Σ applied in this paper were first introduced by Andersson and Perlman (1988), hereafter abbreviated as [AP]. The basic identities (3.14) and (4.17) in [AP] will be applied in Section 3 to obtain the fundamental factorization (3.12) of the likelihood function of the general lattice-ordered CI model for missing data. Although the mathematical derivations in [AP] will not be repeated, some of the essential concepts and notation regarding finite distributive lattices will be reviewed here, along with several examples illustrating the applications of these concepts to the analysis of multivariate normal missing data models. Nonetheless, some familiarity with Sections 1, 3.3, 3.4, and 5 of [AP] will aid the reader of the present paper.

In Section 2 of this paper the general multivariate normal missing data model is formally introduced and monotone and non-monotone incomplete data patterns formally defined. In Section 3 the lattice-ordered CI model determined by an arbitrary incomplete data pattern $\mathbb{S}$ is defined in terms of the finite distributive lattice $\mathcal{K} = \mathcal{K}(\mathbb{S})$ generated from $\mathbb{S}$ by intersections and unions, and the fundamental factorizations (3.12), (3.15), and (3.16) of its likelihood function and parameter space are obtained. This is then shown to yield explicit conditions for existence and uniqueness of MLE under the CI model and explicit expressions for these MLE. Several examples are presented in Section 4 to illustrate the general theory, while some additional comments are given in Section 5. The reader is encouraged to examine the examples in Section 4 as early as possible in order to illuminate the general theory which, although expressed in terms of abstract lattice-theoretic concepts, is actually quite easy to apply to specific incomplete data patterns.

2. The general multivariate normal missing data model.

Let I be a finite index set with $|I| = p$, where $|A|$ denotes the number of elements in the set A. Let $N(\mu,\Sigma)$ denote the multivariate normal distribution on $\mathbb{R}^I$ with mean $\mu \in \mathbb{R}^I$ and covariance $\Sigma \in \mathbb{P}(I)$, the set of positive definite $I\times I$ matrices. Let $y = (x_1,\cdots,x_n) \in \mathbb{M}(I\times N)$ be a collection of independent random column vectors with each $x_j \sim N(\mu,\Sigma)$, where $N = \{1,\cdots,n\}$ and $\mathbb{M}(I\times N)$ denotes the vector space of all real $I\times N$ matrices. The general multivariate missing data model can be described as follows.

Let $\mathfrak{D}(I)$ denote the ring of all subsets of I. For each $j \in N$ let $K_j \in \mathfrak{D}(I)$ denote that subset of I such that the K_j-subvector of x_j is observed while the $I\backslash K_j$-subvector of x_j is missing. To avoid trivialities it is assumed that $K_j \neq \emptyset$ and $\cup(K_j \mid j \in N) = I$; these conditions insure that no column (respectively, row) of y is completely missing.

For each $K \in \mathfrak{D}(I)$ define

$$N_K = \{j \in N \mid K_j = K)$$

$$\mathfrak{N} = (N_K \mid K \in \mathfrak{D}(I)).$$

Then $\mathfrak{N}$ is an arbitrary family of disjoint and possibly empty subsets of N such that $N_\emptyset = \emptyset$ and

$$\cup(N_K \mid K \in \mathfrak{D}(I)) = N \tag{2.1}$$

$$\cup(K \mid K \in \mathfrak{D}(I),\ N_K \neq \emptyset) = I. \tag{2.2}$$

For each $K \in \mathfrak{D}(I)$ let $y^K \in \mathbb{M}(K\times N_K)$ denote the $K\times N_K$ submatrix of y. The projection mapping

$$\begin{aligned} \mathbb{M}(I\times N) &\rightarrow E^{\mathfrak{N}} = \times(\mathbb{M}(K\times N_K) \mid K \in \mathfrak{D}(I)) \\ y &\rightarrow y^{\mathfrak{N}} = (y^K \mid K \in \mathfrak{D}(I)) \end{aligned} \tag{2.3}$$

sends the complete data matrix y to the incomplete data array $y^{\mathfrak{N}}$ actually observed, while the remaining entries of y are missing.

For $(\mu,\Sigma) \in \mathbb{R}^I \times \mathbb{P}(I)$, the distribution of $y^{\mathfrak{N}}$ induced by the projection (2.3) is the multivariate normal distribution $N^{\mathfrak{N}}(\mu,\Sigma)$ on $E^{\mathfrak{N}}$ with density function f given by

$$(2.4) \qquad f = \Pi((\det\Sigma_K)^{-n_K/2} \times \exp\{-\mathrm{tr}(\Sigma_K^{-1}(y^K-\mu^K)(y^K-\mu^K)^t)/2\} \mid K \in \mathfrak{D}(I)),$$

where $n_K = |N_K| \geq 0$, $\Sigma_K \in \mathbb{P}(K)$ is the $K \times K$ submatrix of Σ, and $\mu^K \in \mathbb{M}(K \times N_K)$ is the $K \times N_K$ matrix with each column equal to μ_K, the K-subcolumn of μ. The *general multivariate normal missing data model* $\mathfrak{M}(\mathfrak{N})$ with observation space $E^{\mathfrak{N}}$ and parameter space $\mathbb{R}^I \times \mathbb{P}(I)$ is defined to be the family

$$(2.5) \qquad \mathfrak{M}(\mathfrak{N}) = (N^{\mathfrak{N}}(\mu,\Sigma) \mid (\mu,\Sigma) \in \mathbb{R}^I \times \mathbb{P}(I)).$$

2.1. The incomplete data pattern.

The class

$$\mathfrak{S} = \mathfrak{S}(\mathfrak{N}) = \{K \mid K \in \mathfrak{D}(I),\ N_K \neq \emptyset\}$$

of subsets of I specifies the collection of partially observed column vectors that actually occur (with repetition) in $y^{\mathfrak{N}}$. The classes $\mathfrak{S}$ corresponding to the incomplete data patterns in Figure 1.2 are exhibited in (1.1). As additional examples, the classes $\mathfrak{S}$ corresponding to the patterns in Figure 1.6 are

$$(2.6) \qquad \{12,\ 13\}, \qquad \{13,\ 124,\ 125\},$$

while the classes $\mathfrak{S}$ corresponding to the patterns in Figure 1.7 are

$$(2.7) \qquad \{12,\ 13,\ 123\}, \qquad \{13,\ 124,\ 125,\ 1235\}.$$

Thus we may identify $\mathfrak{S}$ with its corresponding pattern and refer to $\mathfrak{S} \equiv \mathfrak{S}(\mathfrak{N})$ as the *incomplete data pattern determined by* $\mathfrak{N}$. Note that condition (2.2) may then be rewritten as

(2.8) $$\cup(K \mid K \in \mathbb{S}) = I.$$

The parameter (μ,Σ) is *identifiable* in the model $\mathfrak{M}(\mathfrak{N})$ if the mapping $(\mu,\Sigma) \to N^{\mathfrak{N}}(\mu,\Sigma)$ from the parameter space $\mathbb{R}^{I} \times \mathbb{P}(I)$ to the set of normal distributions on $E^{\mathfrak{N}}$ is 1-1. Clearly (μ,Σ) is identifiable if $I \in \mathbb{S}$, i.e., whenever at least one column of y is completely observed. More generally, it can be readily seen that (μ,Σ) is identifiable in $\mathfrak{M}(\mathfrak{N})$ if and only if

(2.9) $$\cup(K \times K \mid K \in \mathbb{S}) = I \times I,$$

i.e., if and only if every pair of the p variates represented in $N(\mu,\Sigma)$ occur together in at least one column of the observation matrix.

The incomplete data pattern $\mathbb{S}$ is called *monotone* (= nested, etc.) if $\mathbb{S}$ is totally ordered under inclusion. In this case (2.8) implies that $I \in \mathbb{S}$, hence by (2.9) the parameter (μ,Σ) is identifiable. Furthermore, the necessary and sufficient condition for existence and uniqueness of the MLE of (μ,Σ) is simply $n_I \geq p+1$ (cf. (4.2)), which reduces to the classical condition $n \geq p+1$ when no data are missing. As pointed out in Section 1, the statistical analysis of a normal model with a monotone data pattern reduces to the analysis of several ordinary linear regression models.

Each pattern $\mathbb{S}$ in (2.6) and (2.7) is non-monotone, however, as are the vast majority of incomplete data patterns. The CI models that simplify the analysis of such patterns are described in the following section.

3. The lattice-ordered conditional independence model determined by an incomplete data pattern.

As defined in Section 2, an incomplete data pattern $\mathbb{S} \equiv \mathbb{S}(\mathcal{N})$ is an arbitrary subclass of $\mathcal{D}(I)\backslash\{\varnothing\}$. The pattern $\mathbb{S}$ uniquely determines the ring $\mathcal{K} \equiv \mathcal{K}(\mathbb{S}) \subseteq \mathcal{D}(I)$ defined to be the smallest subring of $\mathcal{D}(I)$ that contains $\mathbb{S}$ and $\varnothing$, i.e., $\mathcal{K}$ is generated from $\mathbb{S}$ and $\varnothing$ by the set operations $\cup$ and $\cap$. Note that under these operations $\mathcal{K}$ is a finite distributive lattice such that $\varnothing$, $I \in \mathcal{K}$ (cf. (2.8)).

The set $\mathbb{P}_{\mathcal{K}}(I) \subseteq \mathbb{P}(I)$ is defined in [AP] as the set of all covariance matrices Σ such that

$$(3.1) \qquad x \sim N(\mu,\Sigma) \;\Rightarrow\; x_L \perp\!\!\!\perp x_M \,|\, x_{L\cap M} \qquad \forall\, L,M \in \mathcal{K},$$

i.e., x_L and x_M are CI given $x_{L\cap M}$, where x_K denotes the K-subcolumn of x for $K \in \mathcal{K}$. If $L\cap M = \varnothing$, (3.1) reduces to $x_L \perp\!\!\!\perp x_M$. Note that (3.1) is ordinarily written in the form

$$(3.2) \qquad x_{L\backslash(L\cap M)} \perp\!\!\!\perp x_{M\backslash(L\cap M)} \,|\, x_{L\cap M} \qquad \forall\, L,M \in \mathcal{K}.$$

Some of these CI conditions are trivially satisfied, e.g., whenever $L \subseteq M$ (cf. Remark 5.1 of [AP]); in particular, if $\mathcal{K}$ is a chain (cf. Example 4.1) then $\mathbb{P}_{\mathcal{K}}(I) = \mathbb{P}(I)$, i.e., Σ is unrestricted. At the other extreme, if $\mathcal{K} = \mathcal{D}(I)$ then under $\mathbb{P}_{\mathcal{K}}(I)$ all components of x are mutually independent i.e., $\Sigma = \mathrm{Diag}(\sigma_{11},\cdots,\sigma_{pp})$.

The *lattice-ordered conditional independence model* $\mathcal{M}^*(\mathcal{N})$ is obtained from $\mathcal{M}(\mathcal{N})$ by restricting the parameter space from $\mathbb{R}^I \times \mathbb{P}(I)$ to $\mathbb{R}^I \times \mathbb{P}_{\mathcal{K}}(I)$, i.e., by imposing the CI restrictions[6] (3.1) $\equiv$ (3.2) on Σ.

3.1. Factorization of the likelihood function.

Because $N_L = \varnothing$ for $L \in \mathcal{D}(I)\backslash\mathcal{K}$, the probability density function (2.4) of $y^{\mathcal{N}}$ may be rewritten as

---

[6] Because $\mathcal{K}(\mathbb{S}) = \mathbb{S}$ when $\mathbb{S}$ is a chain (= totally ordered), and because $\mathbb{P}_{\mathcal{K}}(I) = \mathbb{P}(I)$ whenever $\mathcal{K}$ is a chain, it follows that $\mathcal{M}^*(\mathcal{N}) = \mathcal{M}(\mathcal{N})$ whenever $\mathbb{S} \equiv \mathbb{S}(\mathcal{N})$ is monotone.

$$(3.3)\qquad f = \Pi(|\det\Sigma_L|^{-n_L/2} \mid L \in \mathcal{K})\times\exp\{-\textstyle\sum(\mathrm{tr}(\Sigma_L^{-1}(y^L-\mu^L)(y^L-\mu^L)^t)/2 \mid L \in \mathcal{K})\}.$$

To show that (3.3) can be factored as a product of density functions of normal linear regression models, we shall apply the basic decomposition formulas (3.14) and (4.17) of [AP]. Their application in (3.3) requires that for each $L \in \mathcal{K}$, the matrices Σ_L, y_L, and μ_L be partitioned according to the *join-irreducible* elements $J(\mathcal{K})$ of the lattice $\mathcal{K}$. This partitioning process, introduced in [AP], §3.3, is now reviewed.

For $K \in \mathcal{K}$, $K \neq \varnothing$, define

$$\langle K\rangle := \cup(K' \in \mathcal{K} \mid K' \subset K),$$

$$[K] := K \setminus \langle K\rangle,$$

hence

$$(3.4)\qquad K = \langle K\rangle \,\dot{\cup}\, [K],$$

where $\dot{\cup}$ indicates that the union is disjoint. Let $J(\mathcal{K})$ denote the poset of non-null join-irreducible elements of the finite distributive lattice $\mathcal{K}$ (cf. [AP], §2), i.e.,

$$J(\mathcal{K}) = \{K \in \mathcal{K} \mid K \neq \varnothing,\ \langle K\rangle \subset K\}.$$

$$= \{K \in \mathcal{K} \mid K \neq \varnothing,\ [K] \neq \varnothing\}.$$

By Remarks 2.1 and 2.2 of [AP],

$$I = \dot{\cup}([K] \mid K \in J(\mathcal{K})),$$

which decomposition determines the partitioning

$$(3.5)\qquad x = (x_{[K]} \mid K \in J(\mathcal{K})), \qquad x \in \mathbb{R}^I.$$

For every $K \in J(\mathcal{K})$ partition Σ_K according to (3.4) as

(3.6) $$\Sigma_K = \begin{pmatrix} \Sigma_{<K>} & \Sigma_{<K]} \\ \Sigma_{[K>} & \Sigma_{[K]} \end{pmatrix},$$

where $\Sigma_{<K>}$ is $<K>\times<K>$, $\Sigma_{[K>}$ is $[K]\times<K>$, $\Sigma_{<K]} = (\Sigma_{[K>})^t$, and $\Sigma_{[K]}$ is $[K]\times[K]$. Furthermore, define

$$\Sigma_{[K]\cdot} \equiv \Sigma_{[K]\cdot<K>} := \Sigma_{[K]} - \Sigma_{[K>}\Sigma^{-1}_{<K>}\Sigma_{<K]}$$

and let $\Sigma^{-1}_{[K]\cdot}$ denote $(\Sigma_{[K]\cdot})^{-1}$. Lastly, for $L \in \mathcal{K}$ with $K \subseteq L$ define $y^L_K \in \mathbb{M}(K\times N_L)$ to be the $K\times N_L$ submatrix of y^L, and partition y^L_K according to (3.4) as

$$y^L_K = \begin{pmatrix} y^L_{<K>} \\ y^L_{[K]} \end{pmatrix}.$$

We now apply (4.17) of [AP] with $(I, \mathcal{K}, \Sigma)$ replaced by $(L, \mathcal{K}_L, \Sigma_L)$, where $\mathcal{K}_L$ is the sublattice of $\mathcal{K}$ defined as $\mathcal{K}_L = \{L' \in \mathcal{K} \mid L' \subseteq L\}$. Since

$$J(\mathcal{K}_L) = J(\mathcal{K})\cap\mathcal{K}_L,$$

$$(\Sigma_L)_{[K]\cdot} = \Sigma_{[K]\cdot}, \qquad K \in J(\mathcal{K}_L),$$

it follows from (4.17) of [AP] that

(3.7) $$\begin{aligned} &\Pi(|\det\Sigma_L|^{-n_L/2} \mid L \in \mathcal{K}) \\ &= \Pi(\Pi(|\det\Sigma_{[K]\cdot}|^{-n_L/2} \mid K \in J(\mathcal{K}_L)) \mid L \in \mathcal{K}) \\ &= \Pi(\Pi(|\det\Sigma_{[K]\cdot}|^{-n_L/2} \mid K \in J(\mathcal{K}),\ K \subseteq L) \mid L \in \mathcal{K}) \\ &= \Pi(\Pi(|\det\Sigma_{[K]\cdot}|^{-n_L/2} \mid L \in \mathcal{K},\ L \supseteq K \mid K \in J(\mathcal{K})) \\ &= \Pi(|\det\Sigma_{[K]\cdot}|^{-n^+_K/2} \mid K \in J(\mathcal{K})), \end{aligned}$$

where

(3.8) $$n_K^+ = \Sigma(n_L \,|\, L \in \mathcal{K},\ L \supseteq K), \qquad K \in J(\mathcal{K}).$$

Next, apply (3.14) of [AP] with $(I, \mathcal{K}, \Sigma, x)$ replaced by $(L, \mathcal{K}_L, \Sigma_L, y^L-\mu^L)$ to obtain

(3.9) $$\Sigma(tr(\Sigma_L^{-1}(y^L-\mu^L)(y^L-\mu^L)^t) \,|\, L \in \mathcal{K})$$

$$= \Sigma(\Sigma(tr(\Sigma_{[K]\cdot}^{-1}(y_{[K]}^L-\mu_{[K]}^L-\Sigma_{[K>}\Sigma_{<K>}^{-1}(y_{<K>}^L-\mu_{<K>}^L))(\cdots)^t) \,|\, K \in J(\mathcal{K}),\ K \subseteq L) \,|\, L \in \mathcal{K})$$

$$= \Sigma(\Sigma(tr(\Sigma_{[K]\cdot}^{-1}(y_{[K]}^L-\mu_{[K]}^L-\Sigma_{[K>}\Sigma_{<K>}^{-1}(y_{<K>}^L-\mu_{<K>}^L))(\cdots)^t) \,|\, L \in \mathcal{K},\ L \supseteq K) \,|\, K \in J(\mathcal{K}))$$

$$= \Sigma(tr(\Sigma_{[K]\cdot}^{-1}(\Sigma(y_{[K]}^L-\mu_{[K]}^L-\Sigma_{[K>}\Sigma_{<K>}^{-1}(y_{<K>}^L-\mu_{<K>}^L))(\cdots)^t) \,|\, L \in \mathcal{K},\ L \supseteq K)) \,|\, K \in J(\mathcal{K})).$$

For $K \in J(\mathcal{K})$, define

$$N_K^+ = \dot{\cup}(N_L \,|\, L \in \mathcal{K},\ L \supseteq K),$$

let $y_K^+ \in \mathcal{M}(K \times N_K^+)$ be the matrix whose $K \times N_L$ submatrix is y_K^L for $L \in \mathcal{K}$, $L \supseteq K$, and partition y_K^+ according to (3.4) as

(3.10) $$y_K^+ = \begin{pmatrix} y_{<K>}^+ \\ y_{[K]}^+ \end{pmatrix}.$$

Then the final expression in (3.9) can be rewritten as

(3.11) $$\Sigma(tr(\Sigma_{[K]\cdot}^{-1}(y_{[K]}^+-\mu_{[K]}^+-\Sigma_{[K>}\Sigma_{<K>}^{-1}(y_{<K>}^+-\mu_{<K>}^+))(\cdots)^t) \,|\, K \in J(\mathcal{K})).$$

By combining (3.7) and (3.11) we conclude that *the density function f given by* (3.3) *of the CI model* $\mathfrak{M}^*(\mathfrak{N})$ *has the following fundamental factorization:*

(3.12)

$$f = \Pi(|\det\Sigma_{[K]\cdot}|^{-n_K^+/2}\times\exp\{-\mathrm{tr}(\Sigma_{[K]\cdot}^{-1}(y_{[K]}^+ - \mu_{[K]}^+ - \Sigma_{[K>}\Sigma_{<K>}^{-1}(y_{<K>}^+ - \mu_{<K>}^+))(\cdots)^t)/2 \mid K\varepsilon J(\mathcal{K})).$$

The K-th factor in (3.12) is the conditional density of $y_{[K]}^+$ given $y_{<K>}^+$, from which it is seen that

$$(3.13) \qquad \begin{aligned} y_{[K]}^+ \mid y_{<K>}^+ &\sim N(\mu_{[K]}^+ + \Sigma_{[K>}\Sigma_{<K>}^{-1}(y_{<K>}^+ - \mu_{<K>}^+),\ \mathrm{Diag}(\Sigma_{[K]\cdot})) \\ &\equiv N(\xi_K^+ + R_K y_{<K>}^+,\ \mathrm{Diag}(\Lambda_K)), \end{aligned}$$

where

$$(3.14) \qquad \begin{aligned} \xi_K^+ &= \mu_{[K]}^+ - \Sigma_{[K>}\Sigma_{<K>}^{-1}\mu_{<K>}^+ \\ R_K &= \Sigma_{[K>}\Sigma_{<K>}^{-1} \\ \Lambda_K &= \Sigma_{[K]} \end{aligned} \quad .$$

Thus *the K-th factor in* (3.12) *is the likelihood function of a multivariate normal linear regression model with regression parameters* ξ_K, R_K *and covariance matrix* Λ_K, *where*

$$\xi_K = \mu_{[K]} - \Sigma_{[K>}\Sigma_{<K>}^{-1}\mu_{<K>} \quad .$$

If we let $f([K] \mid <K>)$ denote the K-th factor in the density function f given by (3.12), then (3.12) assumes the abbreviated form

$$(3.15) \qquad f = \Pi(f([K] \mid <K>) \mid K \ \varepsilon\ J(\mathcal{K})).$$

Since [K] = K when <K> = ∅ we write $f(K)$ for $f([K] \mid \varnothing)$. Equations (1.2) - (1.5) are special cases of (3.15).

3.2. Factorization of the parameter space.

The parameters (ξ_K, R_K, Λ_K), $K \in J(\mathcal{K})$, are called the *$\mathcal{K}$-parameters* of the CI missing data model $\mathfrak{M}^*(\mathfrak{N})$ (cf. [AP], §3.3)[7]. By means of the algorithm for reconstructing (μ,Σ) from its $\mathcal{K}$-parameters presented below, the mapping

$$(3.16) \qquad \mathbb{R}^{I}\times\mathbb{P}_{\mathcal{K}}(I) \rightarrow \times(\mathbb{R}^{[K]}\times\mathbb{M}([K]\times\langle K\rangle)\times\mathbb{P}([K]) \mid K \in J(\mathcal{K}))$$

$$(\mu,\Sigma) \rightarrow ((\xi_K, R_K, \Lambda_K) \mid K \in J(\mathcal{K}))$$

can be shown to be a 1-1 correspondence, so *the parameter space of the model* $\mathfrak{M}^*(\mathfrak{N})$ *is thereby represented as the product of the parameter spaces of the linear regression models given by* (3.13). In summary, it follows from (3.12), (3.13), and (3.16) that *the analysis of the CI missing data model can be reduced to the analysis of* $q \equiv |J(\mathcal{K})|$ *multivariate linear regression models,* as in the case of a monotone pattern. From this it is seen that the $\mathcal{K}$-parameters of (μ,Σ) are identifiable, hence (μ,Σ) is identifiable under the restriction $\Sigma \in \mathbb{P}_{\mathcal{K}}(I)$.

3.3. The reconstruction algorithm.

We now describe the process of reconstructing (μ,Σ) from its $\mathcal{K}$-parameters $((\xi_K, R_K, \Lambda_K) \mid K \in J(\mathcal{K}))$. Under the CI model $\mathfrak{M}^*(\mathfrak{N})$ the MLE $(\hat{\mu}, \hat{\Sigma})$ is obtained by first finding the MLE $((\hat{\xi}_K, \hat{R}_K, \hat{\Lambda}_K) \mid K \in J(\mathcal{K}))$ of the $\mathcal{K}$-parameters, then applying the reconstruction algorithm to obtain $(\hat{\mu}, \hat{\Sigma})$.

The reconstruction algorithm is a direct extension of the stepwise algorithm described in Remark 3.6 of [AP] for reconstructing Σ from its $\mathcal{K}$-parameters $((R_K, \Lambda_K) \mid K \in J(\mathcal{K}))$. Simply follow Remark 3.6 of [AP] with the following changes:

(i) Replace (3.19) of [AP] by the list $((\xi_k, R_k, \Lambda_k) \mid k=1,\cdots,q)$, where $q = |J(\mathcal{K})|$ and K_k is abbreviated by k as in [AP].

(ii) Modify Steps $1, 2, 3,\cdots, k$ in Remark 3.6 of [AP] as follows (Step 3b is unchanged):

---

[7]Recall that $\mathcal{K}$ is uniquely determined by $\mathbb{S} \equiv \mathbb{S}(\mathfrak{N})$. However, different patterns $\mathbb{S}$ may determine the same lattice $\mathcal{K}$, cf. the Examples in Section 4.

Step 1*:

$$\Sigma_1 = \Lambda_1$$
$$\mu_1 = \xi_1$$

Step 2*:

$$\Sigma_{[2>} = R_2\Sigma_1$$
$$\Sigma_{[2]} = \Lambda_2 + R_2\Sigma_{<2]}$$
$$\mu_{[2]} = \xi_2 + R_2\mu_1$$

Step 3*:

$$\Sigma_{[3>} = R_3\Sigma_{<3>}$$
$$\Sigma_{[3]} = \Lambda_3 + R_3\Sigma_{<3]}$$
$$\mu_{[3]} = \xi_3 + R_3\mu_{<3>}$$
$$\Sigma_{[3\ngtr} = R_3\Sigma_{<3\ngtr}$$

$$\vdots$$

Step k*:

$$\Sigma_{[k>} = R_k\Sigma_{<k>}$$
$$\Sigma_{[k]} = \Lambda_k + R_k\Sigma_{<k]}$$
$$\mu_{[k]} = \xi_k + R_k\mu_{<k>}$$
$$\Sigma_{[k\ngtr} = R_k\Sigma_{<k\ngtr} \quad .$$

(iii) In the discussion accompanying these steps in [AP] replace expressions of the form I_K by K, $K \in \mathcal{K}$, and I_k by k, $k = 1, \cdots, q$, and replace the symbol $\vee$ by $\cup$. In the paragraph following Step 2 in [AP], insert "and the subvector $\mu_{1\cup 2}$ (= μ_2 here)" after "the submatrix $\Sigma_{1\cup 2}$ (= Σ_2 here)", and insert "and $\mu_{<3>}$ is a subvector of $\mu_{1\cup 2}$" after "$\Sigma_{<3>}$ is a submatrix of $\Sigma_{1\cup 2}$". In the new paragraph following Step 3b in [AP], insert "and the subvector $\mu_{1\cup\cdots\cup(k-1)}$" after "the submatrix $\Sigma_{1\cup\cdots\cup(k-1)}$", and insert "and μ_k" after " $\Sigma_{1\cup\cdots\cup k}$". In the paragraph immediately preceding Step k in [AP], insert "and $\mu_{<k>}$ is a subvector of $\mu_{1\cup\cdots\cup(k-1)}$" after "$\Sigma_{1\cup\cdots\cup(k-1)}$". In the final paragraph of Remark 3.6 of [AP], insert "and the subvector $\mu_{1\cup\cdots\cup k}$" after "the submatrix $\Sigma_{1\cup\cdots\cup k}$", and insert "and $\mu_{1\cup\cdots\cup q} \equiv \mu$" after "$\Sigma_{1\cup\cdots\cup q} \equiv \Sigma$".

3.4. The maximum likelihood estimator of (μ,Σ).

By (3.12) and well-known results for the multivariate normal linear regression model, for each $K \in J(\mathcal{K})$ the MLE $(\hat{\xi}_K, \hat{R}_K, \hat{\Lambda}_K)$ exists if and only if $n_K^+ \geq |K|+1$. Let e_K denote the N_K^+-column vector each of whose entries is $1/n_K^+$, define

$$
\text{(3.17)} \qquad
\begin{aligned}
\overline{y_K} &= y_K^+ e_K \\
\overline{y_K^+} &= y_K^+(n_K^+ e_K e_K{}^t) \\
S_K &= (y_K^+ - \overline{y_K^+})(y_K^+ - \overline{y}^+)^t,
\end{aligned}
$$

partition the K×K matrix S_K as in (3.6), and partition the K-column vector $\overline{y_K}$ according to (3.4). Then the MLE $(\hat{\xi}_K, \hat{R}_K, \hat{\Lambda}_K)$ is given by

$$
\text{(3.18)} \qquad
\begin{aligned}
\hat{\xi}_K &= \overline{y_{[K]}} - S_{[K>}S_{<K>}^{-1}\overline{y_{<K>}} \\
\hat{R}_K &= S_{[K>}S_{<K>}^{-1} \\
\hat{\Lambda}_K &= S_{[K]}.
\end{aligned}
$$

In view of the factorizations (3.12) and (3.16), it follows that under the CI model $\mathcal{M}^*(\mathcal{N})$, *the* MLE $(\hat{\mu},\hat{\Sigma})$ *for* $(\mu,\Sigma) \in \mathbb{R}^{I}\times\mathbb{P}_{\mathcal{K}}(I)$ *exists for* a. e. $y \in E^{\mathcal{N}}$ *if and only if*

$$
\text{(3.17)} \qquad n_K^+ \geq |K|+1 \qquad \forall\, K \in J(\mathcal{K}).
$$

Since $|K'| \leq |K|$ and $n_{K'}^+ \geq n_K^+$ whenever $K' \subseteq K$ and $K', K \in J(\mathcal{K})$, the condition (3.17) need be verified only for every *maximal* element K of the poset $J(\mathcal{K})$. When a MLE $(\hat{\mu},\hat{\Sigma})$ exists, it is unique and is explicitly obtained by applying the reconstruction algorithm of Section 3.3 to the family $((\hat{\xi}_K, \hat{R}_K, \hat{\Lambda}_K)\,|\,K \in J(\mathcal{K}))$ given by (3.18).

4. Examples.

Since different incomplete data patterns $\mathcal{S}$ may generate the same distributive lattice $\mathcal{K}$, the family of lattice-ordered CI models $\mathfrak{M}^*(\mathfrak{N})$ for incomplete multivariate data arrays is divided into equivalence classes indexed by the family of all finite distributive lattice diagrams. In Examples 4.1 - 4.11 a lattice $\mathcal{K} \subseteq \mathcal{D}(I)$ is selected, the associated CI covariance restrictions are described, the factorization (3.12) ≡ (3.15) of the likelihood function for the models $\mathfrak{M}^*(\mathfrak{N})$ that give rise to $\mathcal{K}$ is determined, the necessary and sufficient condition (3.17) for the existence of the MLE is specified, and the class of all patterns $\mathcal{S}$ that generate $\mathcal{K}$ is described. In Examples 4.12 and 4.13, specific incomplete data patterns $\mathcal{S}$ that appear in the literature are presented, then the lattices $\mathcal{K} \equiv \mathcal{K}(\mathcal{S})$ are determined and the corresponding missing data models analyzed as above.

Each Example is accompanied by a Figure displaying the lattice diagram for $\mathcal{K}$. In these Figures, the members of the poset $J(\mathcal{K})$ are indicated by open circles while the remaining members of $\mathcal{K}$ are indicated by solid dots. The minimal element $\emptyset$ appears at the left of each diagram while the maximal element I appears at the right. From the Figures, notice that $K \in J(\mathcal{K})$ iff K covers exactly one other element of $\mathcal{K}$, i.e., iff exactly one line connects K with elements to its left in the lattice diagram.

Example 4.1. (Monotone data patterns). If $\mathcal{K} \equiv \mathcal{K}_1$ is an ascending chain, i.e., $\emptyset \equiv K_0 \subset K_1 \subset \cdots \subset K_q \equiv I$ (cf. Figure 4.1) then (3.1) is trivially satisfied and $\mathbb{P}_{\mathcal{K}}(I) = \mathbb{P}(I)$, i.e., no CI restrictions are imposed on Σ (cf. Examples 3.1 and 3.2 of [AP]).

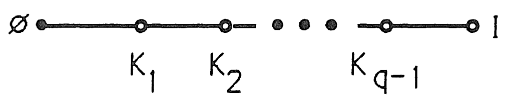

Figure 4.1: The lattice $\mathcal{K}_1$.

Here $J(\mathcal{K}_1) = \{K_1, \cdots, K_q \equiv I\}$ and $\langle K_k \rangle = K_{k-1}$, $k = 1, \cdots, q$. For every missing data model $\mathfrak{M}^*(\mathfrak{N})$ with $\mathcal{K}(\mathcal{S}(\mathfrak{N})) = \mathcal{K}_1$, the fundamental factorization (3.15) of the likelihood function f therefore assumes the form

$$f = f(K_1) f([K_2] | K_1) \cdots f([I] | K_{q-1}). \tag{4.1}$$

Since I is the only maximal element of the poset $J(\mathcal{K}_1)$, condition (3.17) for the existence of the MLE becomes simply

$$n_I \geq p + 1. \tag{4.2}$$

The only data pattern $\mathcal{S}$ that generates $\mathcal{K}_1$ is $\mathcal{S} = J(\mathcal{K}_1)$. In the special case where $I = 12\cdots p$ and $K_k = 12\cdots k$ for $k = 1, \cdots, p \equiv q$, then $[K_k] = \{k\}$ and (4.1) reduces to (1.2). □

The remaining Examples in this Section treat non-monotone incomplete data patterns.

Example 4.2. (Independence of two blocks). Consider the lattice $\mathcal{K} \equiv \mathcal{K}_2$ in Figure 4.2 (cf. Example 3.3 in [AP]):

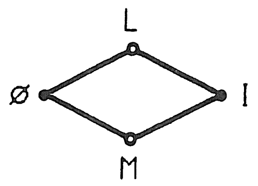

Figure 4.2: The lattice $\mathcal{K}_2$.

Here $J(\mathcal{K}_2) = \{L, M\}$ and $\langle L \rangle = \langle M \rangle = \emptyset$. The partitioning (3.5) and the CI condition (3.2) reduce to

$$x = (x_L, x_M) \tag{4.3}$$

$$x_L \perp\!\!\!\perp x_M, \tag{4.4}$$

respectively, so $\Sigma \in \mathbb{P}_{\mathcal{K}}(I)$ iff $\Sigma = \mathrm{Diag}(\Sigma_L, \Sigma_M)$. The factorization (3.15) becomes

$$f = f(L) f(M), \tag{4.5}$$

and the condition (3.17) for the existence of the MLE becomes

(4.6)
$$n_L^+ \equiv n_L + n_I \geq |L| + 1$$
$$n_M^+ \equiv n_M + n_I \geq |M| + 1,$$

since L and M are the maximal elements of $J(\mathcal{K}_2)$.

In this example, $\mathcal{S} \subseteq \mathcal{D}(I)$ generates $\mathcal{K}_2$ iff $\{L, M\} \subseteq \mathcal{S}$, so there are 2 possible patterns $\mathcal{S}$ such that $\mathcal{K}_2 = \mathcal{K}(\mathcal{S})$:

(4.7)
$$\mathcal{S} = \{L, M, I\}, \qquad \{L, M\}.$$

If I = 12, for example, the patterns

(4.8)
$$\mathcal{S} = \{1, 2, 12\}, \qquad \{1, 2\}$$

(cf. (1.1)) have the forms in (4.7). For both patterns, the CI restriction (4.4) thus reduces to $x_1 \perp\!\!\!\perp x_2$ and the factorization (4.5) reduces to $f = f(1)f(2)$. For the first pattern in (4.8) the MLE existence condition (4.6) becomes

$$n_1 + n_{12} \geq 2, \qquad n_2 + n_{12} \geq 2,$$

while for the second pattern (4.6) reduces to $n_1 \geq 2$, $n_2 \geq 2$.[8]

If I = 123 the patterns

(4.9)
$$\mathcal{S} = \{12, 3, 123\}, \qquad \{12, 3\}$$

(cf. the third patterns in Figures 1.4a and 1.4b) also generate $\mathcal{K}_2$, so (4.4) reduces to $(x_1, x_2) \perp\!\!\!\perp x_3$ and (4.5) reduces to $f = f(12)f(3)$. For the first pattern in (4.9) condition (4.6) becomes

---

[8] More generally, if the variates labelled "1 and "2" actually represent multivariate blocks of variates of dimensions p_1 and p_2, respectively, (cf. Footnote 1) then condition (4.6) becomes $n_1+n_{12} \geq p_1+1$, $n_2+n_{12} \geq p_2+1$ for the first pattern in (4.8) and $n_1 \geq p_1+1$, $n_2 \geq p_2+1$ for the second pattern.

$$n_{12} + n_{123} \geq 3, \quad n_3 + n_{123} \geq 2,$$

while for the second pattern (4.6) reduces to $n_{12} \geq 3$, $n_3 \geq 2$.(9) □

Example 4.3. (One pairwise CI condition). Consider the lattice $\mathcal{K} \equiv \mathcal{K}_3$ in Figure 4.3 (cf. Example 3.5 in [AP]):

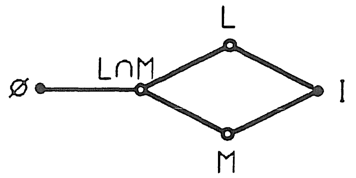

Figure 4.3: The lattice $\mathcal{K}_3$.

Here $J(\mathcal{K}_3) = \{L \cap M, L, M\}$, $\langle L \cap M \rangle = \emptyset$, $\langle L \rangle = \langle M \rangle = L \cap M$. The partitioning (3.5) and the CI condition (3.2) reduce to

(4.10) $$x = (x_{L \cap M}, x_{[L]}, x_{[M]})$$

(4.11) $$x_{[L]} \perp\!\!\!\perp x_{[M]} \mid x_{L \cap M},$$

respectively. The class $\mathbb{P}_{\mathcal{K}}(I)$ is described in (3.37) of [AP]. The factorization (3.15) becomes

(4.12) $$f = f(L \cap M) f([L] \mid L \cap M) f([M] \mid L \cap M),$$

and the MLE existence condition (3.17) is again (4.6).

In this example, $\mathcal{S} \subseteq \mathcal{D}(I)$ generates $\mathcal{K}_3$ iff $\{L, M\} \subseteq \mathcal{S}$, so there are $2^2 = 4$ possible patterns $\mathcal{S}$ such that $\mathcal{K}_3 = \mathcal{K}(\mathcal{S})$:

---

[9] More generally (cf. Footnote 8), condition (4.6) becomes $n_{12}+n_{123} \geq p_1+p_2+1$, $n_3+n_{123} \geq p_3+1$ for the first pattern in (4.9) and $n_{12} \geq p_1+p_2+1$, $n_3 \geq p_3+1$ for the second pattern.

(4.13) $\mathcal{S} = \{L, M\}, \quad \{L\cap M, L, M\}, \quad \{L, M, I\}, \quad \{L\cap M, L, M, I\}.$

If $I = 123$, for example, the patterns

(4.14) $\mathcal{S} = \{12, 13, 123\}, \quad \{12, 13\}$

(cf. the second patterns in Figures 1.4a and 1.4b) have the forms $\{L, M, I\}$ and $\{L, M\}$, respectively. For both patterns, the CI restriction (4.11) thus becomes $x_2 \perp\!\!\!\perp x_3 \mid x_1$ and the factorization (4.12) reduces to (1.4). For the first pattern in (4.14) condition (4.6) becomes

$$n_{12} + n_{123} \geq 3, \quad n_{13} + n_{123} \geq 3,$$

while for the second pattern (4.6) reduces to $n_{12} \geq 3$, $n_{13} \geq 3$. □

Example 4.4. (Marginal independence of two blocks). Consider the lattice $\mathcal{K} \equiv \mathcal{K}_4$ in Figure 4.4 (cf. Example 3.4 of [AP]):

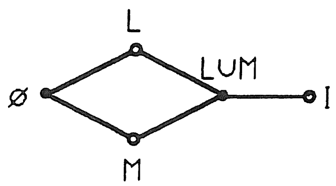

Figure 4.4: The lattice $\mathcal{K}_4$.

Now $J(\mathcal{K}_4) = \{L, M, I\}$, $\langle L\rangle = \langle M\rangle = \emptyset$, and $\langle I\rangle = L\cup M$. The partitioning (3.5) becomes

(4.15) $$x = (x_L, x_M, x_{[I]}),$$

and the CI condition (3.2) again becomes (4.4), so $\Sigma \in \mathbb{P}_{\mathcal{K}}(I)$ iff $\Sigma_{L\cup M} = \mathrm{Diag}(\Sigma_L, \Sigma_M)$. The factorization (3.15) becomes

(4.16) $$f = f(L)f(M)f([I] \mid L\cup M),$$

and the MLE existence condition (3.17) takes the simple form

(4.17) $$n_I \geq p+1$$

since I is the only maximal element of $J(\mathcal{K}_4)$.

In this example $\mathcal{S} \subseteq \mathcal{D}(I)$ generates $\mathcal{K}_4$ iff $\{L, M, I\} \subseteq \mathcal{S}$, so there are 2 possible patterns $\mathcal{S}$ such that $\mathcal{K}_4 = \mathcal{K}(\mathcal{S})$. For example, if $I = 123$ the pattern

(4.18) $$\mathcal{S} = \{1, 2, 123\}$$

(cf. the fifth pattern in Figure 1.4a) has the form $\{L, M, I\}$, hence generates $\mathcal{K}_4$. In this case (4.4) reduces to $x_1 \perp\!\!\!\perp x_2$, (4.16) reduces to $f = f(1)f(2)f(3 \mid 12)$, and (4.17) becomes $n_{123} \geq 4$. □

Example 4.5. (One marginal pairwise CI condition). Consider the lattice $\mathcal{K} \equiv \mathcal{K}_5$ in Figure 4.5 (cf. Example 3.6 of [AP]):

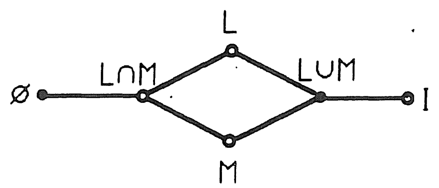

Figure 4.5: The lattice $\mathcal{K}_5$.

Now $J(\mathcal{K}_5) = \{L\cap M, L, M, I\}$, $\langle L\cap M\rangle = \emptyset$, $\langle L\rangle = \langle M\rangle = L\cap M$, and $\langle I\rangle = L\cup M$. The partitioning (3.5) becomes

(4.19) $$x = (x_{L\cap M}, x_{[L]}, x_{[M]}, x_{[I]})$$

and the CI condition (3.2) becomes (4.11). (The class $\mathbf{P}_{\mathcal{K}}(I)$ is described in (3.42) of [AP].) The factorization (3.15) becomes

(4.20) $$f = f(L\cap M)f([L] \mid L\cap M)f([M] \mid L\cap M)f([I] \mid L\cup M)$$

and the condition (3.17) again takes the simple form (4.17).

In this example $\mathcal{S} \subseteq \mathcal{D}(I)$ generates $\mathcal{K}_5$ iff $\{L, M, I\} \subseteq \mathcal{S}$, so there are $2^2 = 4$ possible patterns $\mathcal{S}$ such that $\mathcal{K}_5 = \mathcal{K}(\mathcal{S})$. If $I = 1234$, for example, the 2 patterns

(4.21) $$\mathcal{S} = \{12, 13, 1234\}, \qquad \{12, 13, 123, 1234\}$$

have the forms $\{L, M, I\}$ and $\{L, M, L\cup M, I\}$, respectively, hence both generate $\mathcal{K}_5$. Here (4.11) becomes $x_2 \perp\!\!\!\perp x_3 \mid x_1$, (4.20) becomes $f = f(1)f(2\mid 1)f(3\mid 1)f(4\mid 123)$, and (4.17) becomes $n_{1234} \geq 5$. $\square$

Example 4.6. (Two pairwise CI conditions). Consider the lattice $\mathcal{K}_6$ in Figure 4.6 (cf. Example 3.7 of [AP]):

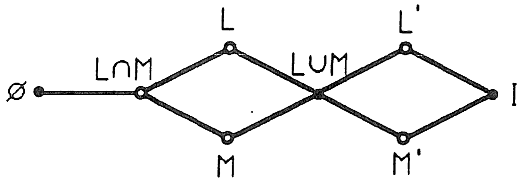

Figure 4.6: The lattice $\mathcal{K}_6$.

Now $J(\mathcal{K}_6) = \{L\cap M, L, M, L', M'\}$, $\langle L\cap M\rangle = \emptyset$, $\langle L\rangle = \langle M\rangle = L\cap M$, and $\langle L'\rangle = \langle M'\rangle = L\cup M$. The partitioning (3.5) and the CI conditions (3.2) assume the respective forms

(4.22) $$x = (x_{L\cap M}, x_{[L]}, x_{[M]}, x_{[L']}, x_{[M']}),$$

(4.23) $$x_{[L]} \perp\!\!\!\perp x_{[M]} \mid x_{L\cap M}, \qquad x_{[L']} \perp\!\!\!\perp x_{[M']} \mid x_{L\cup M}.$$

(The class $\mathbb{P}_{\mathcal{K}}(I)$ is described in Example 3.7 of [AP].) The factorization (3.15) and the MLE existence condition (3.17) become, respectively,

(4.24) $$f = f(L\cap M)f([L]\mid L\cap M)f([M]\mid L\cap M)f([L']\mid L\cup M)f([M']\mid L\cup M),$$

(4.25) $$\begin{aligned} n^{+}_{L'} &\equiv n_{L'} + n_I \geq |L'| + 1 \\ n^{+}_{M'} &\equiv n_{M'} + n_I \geq |M'| + 1, \end{aligned}$$

since L' and M' are the only maximal elements of $J(\mathcal{K}_6)$.

A pattern $\mathcal{S} \subseteq \mathcal{D}(I)$ generates $\mathcal{K}_6$ iff $\{L, M, L', M'\} \subseteq \mathcal{S}$, so there are $2^3 = 8$ possible patterns $\mathcal{S}$ such that $\mathcal{K}_6 = \mathcal{K}(\mathcal{S})$. If $I = 12345$ two such patterns are

(4.26) $\mathcal{S} = \{12, 13, 123, 1234, 1235\}, \quad \{12, 13, 1234, 1235, 12345\},$

which have the forms $\{L, M, L \cup M, L', M'\}$ and $\{L, M, L', M', I\}$, respectively. For these patterns, (4.23) and (4.24) become

$$x_2 \perp\!\!\!\perp x_3 \mid x_1, \qquad x_4 \perp\!\!\!\perp x_5 \mid (x_1, x_2, x_3),$$

$$f = f(1)f(2 \mid 1)f(3 \mid 1)f(4 \mid 123)f(5 \mid 123),$$

respectively, while for the first pattern (4.25) becomes

$$n_{1234} \geq 5, \qquad n_{1235} \geq 5$$

and for the second pattern (4.25) becomes

$$n_{1234} + n_{12345} \geq 5, \qquad n_{1235} + n_{12345} \geq 5. \quad \square$$

Example 4.7. (Independence of three blocks). Consider the lattice $\mathcal{K}_7$ in Figure 4.7:

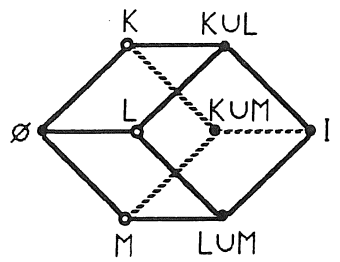

Figure 4.7: The lattice $\mathcal{K}_7$.

Unlike the preceeding examples, $\mathcal{K}_7$ is a non-planar lattice. Here $J(\mathcal{K}_7) = \{K, L, M\}$, and $\langle K \rangle = \langle L \rangle = \langle M \rangle = \varnothing$. The partitioning (3.5) assumes the form

$$x = (x_K, x_L, x_M), \tag{4.27}$$

while (3.2) and (3.15) reduce to

$$x_K \perp\!\!\!\perp x_L \perp\!\!\!\perp x_M, \tag{4.28}$$

$$f = f(K)f(L)f(M), \tag{4.29}$$

respectively, so $\Sigma \in \mathbb{P}_{\mathcal{K}}(I)$ iff $\Sigma = \mathrm{Diag}(\Sigma_K, \Sigma_L, \Sigma_M)$. . Since K, L, M are the maximal elements of $J(\mathcal{K}_7)$, (3.17) becomes

$$\begin{aligned} n_K^+ &\equiv n_K + n_{K\cup L} + n_{K\cup M} + n_I \geq |K| + 1 \\ n_L^+ &\equiv n_L + n_{K\cup L} + n_{L\cup M} + n_I \geq |L| + 1 \\ n_M^+ &\equiv n_M + n_{K\cup M} + n_{L\cup M} + n_I \geq |M| + 1. \end{aligned} \tag{4.30}$$

A pattern $\mathcal{S} \subseteq \mathcal{D}(I)$ generates $\mathcal{K}_7$ iff $\mathcal{S}$ contains one of the following five subpatterns: {K, L, M}, {K, L, K∪M, L∪M}, {K, M, K∪L, L∪M}, {L, M, K∪L, K∪M}, {K∪L, K∪M, L∪M}; there are 36 such patterns. If I = 123, for example, two such patterns are

$$\mathcal{S} = \{12, 13, 23, 123\}, \quad \{12, 13, 23\} \tag{4.31}$$

(cf. the first patterns in Figures 1.4a and 1.4b), which have the forms {K∪L, K∪M, L∪M, I} and {K∪L, K∪M, L∪M}, respectively, with K = 1, L = 2, M = 3. For these two patterns (4.28) becomes $x_1 \perp\!\!\!\perp x_2 \perp\!\!\!\perp x_3$ while (4.30) reduces to

$$n_{12} + n_I \geq 3, \quad n_{13} + n_I \geq 3, \quad n_{23} + n_I \geq 3. \quad \square$$

Example 4.8. (Conditional independence of three blocks). Consider the lattice $\mathcal{K}_8$ in Figure 4.8:

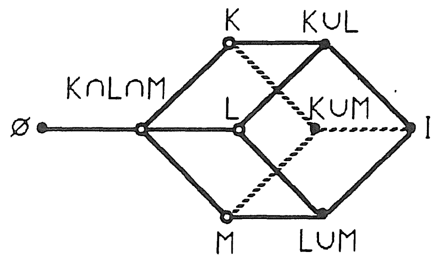

Figure 4.8: The lattice $\mathcal{K}_8$.

Here $J(\mathcal{K}_8) = \{K\cap L\cap M, K, L, M\}$, while $\langle K\cap L\cap M\rangle = \emptyset$, $\langle K\rangle = \langle L\rangle = \langle M\rangle = K\cap L\cap M$. The partitioning (3.5) assumes the form

(4.32) $$x = (x_{K\cap L\cap M}, x_{[K]}, x_{[L]}, x_{[M]}),$$

while (3.2) and (3.15) reduce to

(4.33) $$x_{[K]} \perp\!\!\!\perp x_{[L]} \perp\!\!\!\perp x_{[M]} \mid x_{K\cap L\cap M},$$

(4.34) $$f = f(K\cap L\cap M) f([K] \mid K\cap L\cap M) f([L] \mid K\cap L\cap M) f([M] \mid K\cap L\cap M),$$

respectively, and (3.17) becomes (4.30).

Again a pattern $\mathcal{S} \subseteq \mathcal{D}(I)$ generates $\mathcal{K}_8$ iff $\mathcal{S}$ contains one of the following five subpatterns: $\{K, L, M\}$, $\{K, L, K\cup M, L\cup M\}$, $\{K, M, K\cup L, L\cup M\}$, $\{L, M, K\cup L, K\cup M\}$, $\{K\cup L, K\cup M, L\cup M\}$; there are 72 such subpatterns.

The incomplete data pattern in Figure 6.3 of Little and Rubin (1987) has the form

$$\mathcal{S} = \{\alpha 1, \alpha 2, \alpha 3\},$$

where $\alpha\cap\{1, 2, 3\} = \emptyset$. It is readily seen that $\mathcal{S}$ generates the lattice $\mathcal{K}_8$ with $K = \alpha 1$, $L = \alpha 2$, $M = \alpha 3$. In this case the partitioning (4.32), the CI restriction (4.33), the factorization (4.34), and the MLE existence condition (4.30) take the respective forms

$$x = (x_\alpha, x_1, x_2, x_3)$$

$$x_1 \perp\!\!\!\perp x_2 \perp\!\!\!\perp x_3 \mid x_\alpha$$

$$f = f(\alpha) f(1 \mid \alpha) f(2 \mid \alpha) f(3 \mid \alpha).$$

$$n_{\alpha 1} \geq |\alpha| + 2, \qquad n_{\alpha 2} \geq |\alpha| + 2, \qquad n_{\alpha 3} \geq |\alpha| + 2. \quad \square$$

Example 4.9. (Independence of two blocks). Consider the lattice $\mathcal{K}_9$ in Figure 4.9:

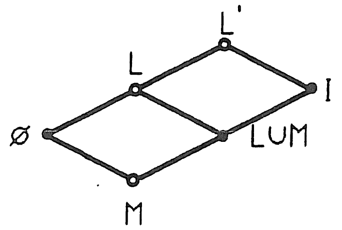

Figure 4.9: The lattice $\mathcal{K}_9$.

Here $J(\mathcal{K}_9) = \{L, M, L'\}$, $\langle L \rangle = \langle M \rangle = \emptyset$, and $\langle L' \rangle = L$. The partitioning (3.5) and the CI condition (3.2) become, respectively

(4.35) $$x = (x_L, x_M, x_{[L']}),$$

(4.36) $$(x_L, x_{[L']}) \perp\!\!\!\perp x_M,$$

a single independence condition. The factorization (3.15) and MLE existence condition (3.17) become

(4.37) $$f = f(L) f(M) f([L'] \mid L),$$

(4.38) $$\begin{aligned} n_M^+ &\equiv n_M + n_{L\cup M} + n_I \geq |M| + 1 \\ n_{L'}^+ &\equiv \qquad\qquad n_{L'} + n_I \geq |L'| + 1, \end{aligned}$$

respectively, since M and L' are the maximal elements of $J(\mathcal{K}_9)$.

A pattern $\mathcal{S} \subseteq \mathcal{D}(I)$ generates $\mathcal{K}_9$ iff $\{M, L', L\} \subseteq \mathcal{S}$ or $\{M, L', L \cup M\} \subseteq \mathcal{S}$: there are $3 \times 2 = 6$ such patterns. If $I = 123$, two such patterns are

$$\mathcal{S} = \{1, 3, 12, 123\}, \quad \{1, 3, 12\} \tag{4.39}$$

(cf. the fourth patterns in Figures 1.4a and 1.4b), which have the forms $\{L, M, L', I\}$ and $\{L, M, L'\}$, respectively. For these patterns, (4.36) becomes $(x_1, x_2) \perp\!\!\!\perp x_3$, (4.37) becomes $f = f(1)f(2 \mid 1)f(3)$, while (4.38) becomes

$$n_3 + n_{123} \geq 2, \quad n_{12} + n_{123} \geq 3$$

for the first pattern and $n_3 \geq 2$, $n_{12} \geq 3$ for the second pattern. The pattern

$$\mathcal{S} = \{3, 12, 13, 123\} \tag{4.40}$$

has the form $\{M, L', L \cup M, I\}$, hence also generates $\mathcal{K}_9$; in this case (4.38) assumes the form

$$n_3 + n_{13} + n_{123} \geq 2, \quad n_{12} + n_{123} \geq 3. \quad \square$$

Example 4.10. (One pairwise CI condition). Consider the lattice $\mathcal{K}_{10}$ in Figure 4.10:

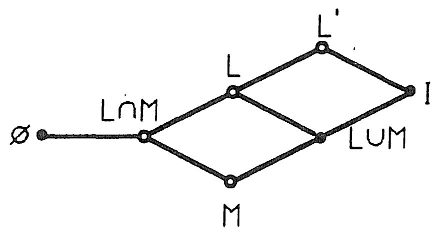

Figure 4.10: The lattice $\mathcal{K}_{10}$.

Here $J(\mathcal{K}_{10}) = \{L \cap M, L, M, L'\}$, $\langle L \cap M \rangle = \emptyset$, $\langle L \rangle = \langle M \rangle = L \cap M$, and $\langle L' \rangle = L$. The partitioning (3.5) and CI condition (3.2) become, respectively,

$$x = (x_{L \cap M}, x_{[L]}, x_{[M]}, x_{[L']}), \tag{4.41}$$

$$(x_{[L]}, x_{[L']}) \perp\!\!\!\perp x_{[M]} \mid x_{L \cap M}. \tag{4.42}$$

a single CI condition. The factorization (3.15) becomes

$$f = f(L\cap M)f([L] \mid L\cap M)f([M] \mid L\cap M)f([L'] \mid L), \tag{4.43}$$

while (3.17) again becomes (4.38).

It is again seen that a pattern $\mathbf{S} \subseteq \mathcal{D}(I)$ generates $\mathcal{K}_{10}$ iff $\{M, L', L\} \subseteq \mathbf{S}$ or $\{M, L', L\cup M\} \subseteq \mathbf{S}$; there are $3\times2^2 = 12$ such patterns.

Rubin (1987, Table 5.6, p. 190), considered the following incomplete data pattern[4] where $I = 1234$:

$$\mathbf{S} = \{1,\ 12,\ 13,\ 123,\ 124,\ 1234\}. \tag{4.44}$$

Clearly $\mathbf{S}$ has the form $\{L\cap M, L, M, L\cup M, L', I\}$, i.e., $\mathbf{S} = \mathcal{K}_{10}\backslash\{\varnothing\}$, hence $\mathbf{S}$ generates $\mathcal{K}_{10}$. Since $L\cap M = 1$, (4.42), (4.43), and (4.38) become, respectively,

$$(x_2, x_4) \perp\!\!\!\perp x_3 \mid x_1\ , \tag{4.45}$$

$$f = f(1)f(2 \mid 1)f(3 \mid 1)f(4 \mid 12), \tag{4.46}$$

$$n_{13} + n_{123} + n_{1234} \geq 3, \qquad n_{124} + n_{1234} \geq 4.$$

Condition (4.45) is the minimal CI assumption under which the analysis of a model $\mathfrak{M}^*(\mathfrak{N})$ with the incomplete data pattern $\mathbf{S}$ in (4.44) can be reduced to the analysis of ordinary linear regression models. However, Rubin (1987) did not discuss the condition (4.45) or the factorization (4.46). Instead, he remarked (bottom of p. 190) that if the data pattern (4.44) were reduced to

$$\mathbf{S}' = \{1,\ 12,\ 123,\ 124,\ 1234\} \tag{4.47}$$

(cf. Figure 1.5) by discarding the observations on variate 3 in block 13 of $\mathbf{S}$, then under the CI assumption

$$x_3 \perp\!\!\!\perp x_4 \mid (x_1, x_2) \tag{4.48}$$

the factorization (1.3) obtains.

To relate this to our general theory, note that the data pattern $\mathbb{S}'$ in (4.47) generates the sublattice $\mathcal{K}'_{10}$ in Figure 4.10a:

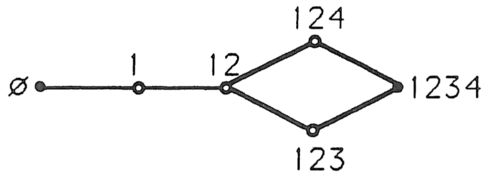

Figure 4.10a: The sublattice $\mathcal{K}'_{10}$.

In this case the CI restriction (3.2) reduces to (4.48), the general factorization (3.15) reduces to (1.3), while (3.17) becomes

$$n_{123} + n_{1234} \geq 4, \qquad n_{124} + n_{1234} \geq 4.$$

Actually, since (4.45) and (4.48) each entail only a single CI restriction, the analysis of the pattern $\mathbb{S}'$ is no simpler than that of $\mathbb{S}$. Because $\mathcal{K}'_{10}$ is a sublattice of $\mathcal{K}_{10}$, however, condition (4.48) imposes fewer restrictions on Σ than does (4.45).

By examining Figure 4.10a it is also seen that a monotone incomplete data pattern $\mathbb{S}''$ can be obtained from $\mathbb{S}'$ either by discarding the observations on variate 3 in block 123 or by discarding the observations on variate 4 in block 12; under $\mathbb{S}''$ *no* restrictions are imposed on Σ. Of course, as suggested in Section 1, loss of estimating efficiency might offset the benefit of fewer covariance restrictions. □

Remark. Of course, it is not usually the case that a *less* restrictive CI covariance model is obtained when some observations are discarded from an incomplete data pattern. For example, if the observations on variate 2 in block 12 of the pattern $\mathbb{S}$ in (4.44) are discarded, then the resulting pattern {1, 13, 123, 124, 1234} generates exactly the same lattice $\mathcal{K}_{10}$ as did $\mathbb{S}$, and hence the same CI covariance model. In fact, if instead the observations on variate 1 in blocks 1 and 12 of $\mathbb{S}$ are discarded, then the resulting pattern {2, 13, 123, 124, 1234} generates a lattice (similar to the lattice $\mathcal{K}_{11}$ below except that $L \cap M = \varnothing$) which is strictly larger than $\mathcal{K}_{10}$ and which determines a *more* restrictive CI covariance model. □

Example 4.11. (Two pairwise CI restrictions). Consider the lattice $\mathcal{K}_{11}$ in Figure 4.11 (cf. Example 3.8 of [AP]):

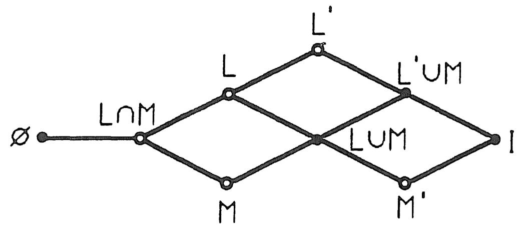

Figure 4.11: The lattice $\mathcal{K}_{11}$.

Here $J(\mathcal{K}_{11}) = \{L\cap M, L, M, L', M'\}$, while $\langle L\cap M\rangle$, $\langle L\rangle$, $\langle M\rangle$, $\langle L'\rangle$ are as in Example 4.10 and $\langle M'\rangle = L\cup M$. The partitioning (3.5) assumes the form (4.22) , while it can be shown that (3.2) and (3.15) are equivalent to

$$(4.49) \qquad (x_{[L]}, x_{[L']}) \perp\!\!\!\perp x_{[M]} \mid x_{L\cap M}, \qquad x_{[L']} \perp\!\!\!\perp (x_{[M]}, x_{[M']}) \mid x_{[L]},$$

$$(4.50) \qquad f = f(L\cap M) f([L] \mid L\cap M) f([M] \mid L\cap M) f([L'] \mid L) f([M'] \mid L\cup M),$$

respectively. (See Remark 5.1 of [AP] for other sets of CI conditions equivalent to (4.49).) Since L' and M' are the maximal elements of $J(\mathcal{K}_{11})$, condition (3.17) for existence of the MLE becomes

$$(4.51) \qquad \begin{aligned} n^+_{L'} &\equiv n_{L'} + n_{L'\cup M} + n_I \geq |L'| + 1 \\ n^+_{M'} &\equiv n_{M'} + n_I \geq |M'| + 1. \end{aligned}$$

A pattern $\mathcal{S} \subseteq \mathcal{D}(I)$ generates $\mathcal{K}_{11}$ iff $\{M, L', M'\} \subseteq \mathcal{S}$. There are $2^5 = 32$ such patterns.

If the lattice $\mathcal{K}_{11}$ is extended to the lattice $\mathcal{K}'_{11}$ in Figure 4.11a, a simpler but more restrictive CI model is obtained (cf. Example 3.9 of [AP]).

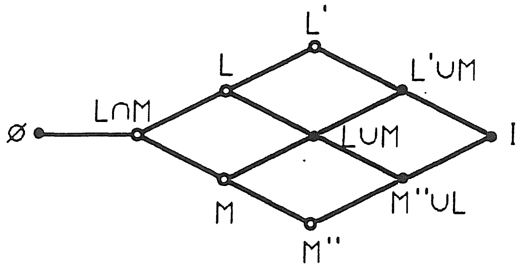

Figure 4.11a: The lattice $\mathcal{K}'_{11}$.

For the CI model determined by $\mathcal{K}'_{11}$ the partitioning (3.5) again assumes the form (4.22) (with $x_{[M']}$ replaced by $x_{[M'']}$), but (4.49) is now replaced by the single CI condition

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

while (4.50) and (4.51) are replaced by

(4.53) $$f = f(L\cap M)\, f([L] \mid L\cap M)\, f([M] \mid L\cap M)\, f([L'] \mid L)\, f([M''] \mid M),$$

(4.54) $$\begin{aligned} n^{+}_{L'} &\equiv n_{L'} + n_{L'\cup M} + n_I \geq |L'| + 1 \\ n^{+}_{M''} &\equiv n_{M''} + n_{L\cup M''} + n_I \geq |M''| + 1. \end{aligned}$$

Even though the CI model determined by $\mathcal{K}'_{11}$ is more restrictive than that determined by $\mathcal{K}_{11}$, its relative simplicity suggests that it might be considered for the analysis of incomplete data patterns that generate $\mathcal{K}_{11}$. □

Example 4.12. Rubin(1974, p. 469, Table 1) considered the incomplete data pattern

$$\mathcal{S} = \{3,\ 8,\ 1238,\ 123678,\ 345678\}, \tag{4.55}$$

where $I = 12345678$. The pattern $\mathcal{S}$ generates the lattice $\mathcal{K}_{12}$ in Figure 4.12:

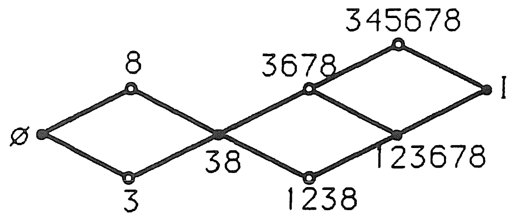

Figure 4.12: The lattice $\mathcal{K}_{12}$.

Since $J(\mathcal{K}_{12}) = \{3,\ 8,\ 1238,\ 3678,\ 345678\}$, the partitioning (3.5) becomes

$$x = (x_3,\ x_8,\ (x_1,\ x_2),\ (x_6,\ x_7),\ (x_4,\ x_5)). \tag{4.56}$$

The CI conditions (3.2), the likelihood factorization (3.15), and the MLE existence condition (3.17) determined by $\mathcal{K}_{12}$ are, respectively,

$$x_3 \perp\!\!\!\perp x_8, \qquad (x_1,\ x_2) \perp\!\!\!\perp (x_4,\ x_5,\ x_6,\ x_7) \mid (x_3,\ x_8), \tag{4.57}$$

$$f = f(3)f(8)f(12 \mid 38)f(67 \mid 38)f(45 \mid 3678), \tag{4.58}$$

$$\begin{aligned} n^{+}_{1238} &\equiv n_{1238} + n_{123678} + n_I \geq 5 \\ n^{+}_{345678} &\equiv n_{345678} + n_I \geq 7, \end{aligned} \tag{4.59}$$

since 1238 and 345678 are the maximal elements of $J(\mathcal{K}_{12})$. For the incomplete data pattern $\mathcal{S}$ in (4.55) considered by Rubin, $n_I = 0$.

From Figure 4.12 it can be seen that the CI model determined by $\mathcal{K}_{12}$ remains applicable if incomplete observations of any of the following forms are added to the pattern $\mathcal{S}$ in (4.55), in which case (4.57), (4.58), and (4.59) remain valid: 38, 3678, 12345678 (= complete). □

Example 4.13. Anderson (1957, eqn. (14)) considered the incomplete data pattern

$$\mathcal{S} = \{13,\ 124,\ 125\}, \tag{4.60}$$

where $I = 12345$ (cf. the second pattern in Figure 1.6). The pattern $\mathcal{S}$ generates the lattice $\mathcal{K}_{13}$ in Figure 4.13:

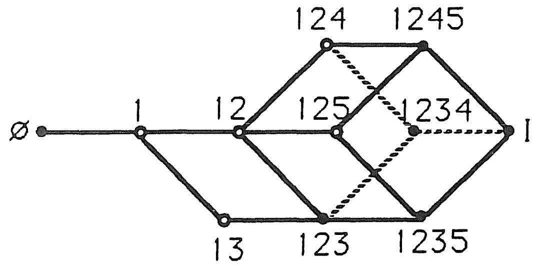

Figure 4.13: The lattice $\mathcal{K}_{13}$.

Here $J(\mathcal{K}_{13}) = \{1,\ 12,\ 13,\ 124,\ 125\}$ and the partitioning (3.5) is simply

$$x = (x_1,\ x_2,\ x_3,\ x_4,\ x_5). \tag{4.61}$$

The CI conditions (3.2), the likelihood factorization (3.15), and the MLE existence condition (3.17) determined by $\mathcal{K}_{13}$ are, respectively,

$$x_2 \perp\!\!\!\perp x_3 \mid x_1, \qquad x_3 \perp\!\!\!\perp x_4 \perp\!\!\!\perp x_5 \mid (x_1,\ x_2), \tag{4.62}$$

$$f = f(1)f(2 \mid 1)f(3 \mid 1)f(4 \mid 12)f(5 \mid 12), \tag{4.63}$$

$$\begin{aligned} n_{13}^{+} &\equiv n_{13} + n_{123} + n_{1234} + n_{1235} + n_I \geq 3 \\ n_{124}^{+} &\equiv n_{124} + n_{1234} + n_{1245} + n_I \geq 4 \\ n_{125}^{+} &\equiv n_{125} + n_{1235} + n_{1245} + n_I \geq 4, \end{aligned} \tag{4.64}$$

since 13, 124, and 125 are the maximal elements of $J(\mathcal{K}_{13})$. For the pattern $\mathcal{S}$ in (4.60) considered by Anderson, only n_{13}, n_{124}, and n_{125} are non-zero.

From Figure 4.13 it can be seen that the CI model determined by $\mathcal{K}_{13}$ remains applicable if incomplete observations of any of the following forms are added to the incomplete data pattern $\mathcal{S}$ in (4.60), in which case (4.62), (4.63), and (4.64) remain valid: 1, 12, 123, 1245, 1235, 12345 (= complete). In particular, the second pattern in Figure 1.7 also generates the lattice $\mathcal{K}_{13}$, hence, as stated at the end of Section 1.3, the above analysis remains applicable. □

5. Comments.

Under a general multivariate normal missing data model $\mathfrak{M}(\mathfrak{N})$ as defined in Section 2, some elements σ_{ij} of Σ may be unidentifiable, hence inestimable. For example, if $I = 123$ and $\mathcal{S} = \{12,13\}$ (cf. (4.14) of Example 4.3) then the covariance σ_{23} (equivalently, the correlation ρ_{23}) is unidentifiable because the variates 2 and 3 are never observed simultaneously. The CI missing data model $\mathfrak{M}^*(\mathfrak{N})$ imposes conditional independence restrictions on Σ under which the unidentifiable covariances are assumed to be functions of the identifiable covariances, which in turn are functions of the $\mathcal{K}$-parameters of Σ. Thus, in the above example the CI restriction $2 \perp\!\!\!\perp 3 \mid 1$ is equivalent to the relation $\sigma_{23} = \sigma_{21}\sigma_{11}^{-1}\sigma_{13}$. Here σ_{11}, σ_{21}, and σ_{13} are functions of the $\mathcal{K}$-parameters σ_{11}, $\sigma_{21}\sigma_{11}^{-1}$, and $\sigma_{31}\sigma_{11}^{-1}$, so once their ML estimates are obtained the ML estimate of σ_{23} is immediately determined. It is important to note that the unidentifiable covariances and correlations are *not* simply set equal to 0 under the CI model determined by $\mathcal{S}$ (unless the CI restrictions are in fact independence restrictions). See Sections 3.3, 3.4, and 5 of [AP] for additional examples.

In order to carry out the likelihood analysis of the missing data model $\mathfrak{M}^*(\mathfrak{N})$, after determining the incomplete data pattern $\mathcal{S}$ it is necessary to determine the poset $J(\mathcal{K})$ of join-irreducible elements of the lattice $\mathcal{K} \equiv \mathcal{K}(\mathcal{S})$ generated by $\mathcal{S}$. This may be carried out in a computationally straightforward manner by first generating $\mathcal{K}$ and then determining $J(\mathcal{K})$. Since $\mathcal{K}$ is distributive, $\mathcal{K}$ may be generated as $\mathcal{K} = \cup(\cap(\mathcal{S}))$, where $\cap(\mathcal{S})$ is the collection of all finite intersections of members of $\mathcal{S}$ and $\cup(\cap(\mathcal{S}))$ is the collection of all finite intersections of members of $\cap(\mathcal{S})$. Then $J(\mathcal{K})$ can be determined from the representation of $\mathcal{K}$ as a directed graph.

In general, however, $\mathcal{K}$ and possibly also $\mathcal{S}$ may be much larger than $J(\mathcal{K})$. For example, if $I = 12\cdots p$ and $\mathcal{S} = \{1, \cdots, p\}$ or $\mathcal{D}(I)\backslash\emptyset$, then $\mathcal{K} = \mathcal{D}(I)$ so $|\mathcal{K}| = 2^p$, while $J(\mathcal{K}) = \{1, \cdots, p\}$ so $|J(\mathcal{K})| = p$. In fact, for every lattice $\mathcal{K} \subseteq \mathcal{D}(I)$ it is true that $|J(\mathcal{K})| \leq p$ (cf. Graetzer (1977, II.1, Corollary 14)). Thus it would be desirable to find a polynomial time(p) algorithm that determines $J(\mathcal{K})$ directly from $\mathcal{S}$ without first generating $\mathcal{K}$, if such an algorithm exists.[10]

Another combinatorial question of a more theoretical nature is the following. Suppose that observations are missing at random from a complete $p \times n$ data matrix y (cf. Section 2) according to a Bernoulli process. As $p, n \to \infty$ at appropriate rates, what is the limiting probability that the resulting incomplete data pattern $\mathcal{S}$ is monotone? Other variations of this question can be readily formulated.

Finally, it is important to note that the results in this paper can be expressed in a coordinate-free way, thus allowing their application to generalized missing data models where some observations may be linear combinations of the original variates. As a simple example, if $x \equiv (x_1, x_2, x_3)$ denotes a complete observation ($p = 3$), then for some individuals in the sample only $x_1+x_2+x_3$ might be observed. Furthermore, the results in this paper can be extended to more general multivariate linear models with missing data, e.g., MANOVA and GMANOVA, and results can be obtained for testing appropriate linear hypotheses as well as for estimating papameters. These topics will be treated in two forthcoming papers by Andersson, Marden, and Perlman (1989a,b) that treat invariant multivariate linear models with monotone and non-monotone missing data patterns, respectively.

Acknowledgement. We gratefully thank Donald Rubin and Richard Ladner for many helpful suggestions and David Perlman for preparing the graphic illustrations.

[10]There is no general inclusion relation between $\mathcal{S}$ and $|J(\mathcal{K})|$. Simple examples can be constructed where $\mathcal{S} = |J(\mathcal{K})|$, $\mathcal{S} \subset |J(\mathcal{K})|$, $\mathcal{S} \supset |J(\mathcal{K})|$, or where none of these relations hold. It is true that $\mathcal{S} \equiv J(\mathcal{K})$ always generates $J(\mathcal{K})$, i.e., $\mathcal{K}(J(\mathcal{K})) = \mathcal{K}$ (cf.Graetzer (1978, II.1, Corollary 13)).

REFERENCES

Anderson, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *J. Amer. Statist. Assoc.* **52**, 200-203.

Andersson, S. A., Marden, J. I., and Perlman, M. D. (1989a). Totally ordered multivariate linear models. In preparation.

Andersson, S. A., Marden, J. I., and Perlman, M. D. (1989b). Partially ordered multivariate linear models. In preparation.

Andersson, S. A. and Perlman, M. D. (1988). Lattice models for conditional independence in a multivariate normal distribution. Submitted for publication.

Bhargava, R. P. (1962). Multivariate tests of hypotheses with incomplete data. Tech. Report No. 3, Applied Mathematics and Statistics Laboratories, Stanford University, Stanford, California.

Bhargava. R. P. (1975). Some one-sample hypothesis testing problems where there is a monotone sample from a multivariate normal population. *Ann. Inst. Statist. Math.* **27**, 327-339.

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

Graetzer, G. (1978). *General Lattice Theory*. Academic Press, New York.

Hocking, R. R. and Smith, W. B. (1968). Estimation of parameters in the multivariate normal distribution with missing observations. *J. Amer. Statist. Assoc.* **63**, 159-173.

Kariya, T., Krishnaiah, P. R., and Rao, C.R. (1983). Inference on parameters of multivariate normal populations when some data is missing. In *Developments in Statistics* (P. R. Krishnaiah, ed.) 4, 137-184, Academic Press, New York.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* John Wiley and Sons, New York.

Lord, F. M. (1955). Estimation of parameters from incomplete data. *J. Amer. Statist. Assoc.* **50**, 870-876.

Murray, G. D. (1977). Discussion of "Maximum likelihood from incomplete data via the EM Algorithm" by A. P. Dempster, N. M. Laird, and D. B. Rubin, *J. Roy. Statist. Soc. Ser. B* **39**, 1-38.

Rao, C. R. (1956). Analysis of dispersion with incomplete observations on one of the characters. *J. Roy. Statist. Soc. Ser. B* **18**, 259-264.

Rubin, D. B. (1974). Characterizing the estimation of parameters in incomplete data problems. *J. Amer. Statist. Assoc.* **69**, 467-474.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Sample Surveys.* John Wiley and Sons, New York.

PREPRINTS 1988

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

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

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

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

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

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

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

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

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

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

PREPRINTS 1989

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

No. 1 Bertelsen, Aksel: Asymptotic Expansion of a Complex Hypergeometric Function.

No. 2 Davidsen, Michael and Jacobsen, Martin: Weak Convergence of Twosided Stochastic Integrals, with an Application to Models for Left Truncated Survival Data.

No. 3 Johansen, Søren: Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models.

No. 4 Johansen, Søren and Juselius, Katarina: The Full Information Maximum Likelihood Procedure for Inference on Cointegration - with Applications.

No. 5 Thompson, Steven K.: Adaptive Cluster Sampling.

No. 6 Thompson, Steven K.: Adaptive Cluster Sampling: Designs with Primary and Secondary Units.

No. 7 Thompson, Steven K.: Stratified Adaptive Cluster Sampling.

No. 8 Johansen, Søren: The Power Function of the Likelihood Ratio Test for Cointegration.

No. 9 Andersson, S.A. and Perlman, M.D.: Lattice-Ordered Conditional Independence Models for Missing Data.