4

Structural Imsets: Fundamentals

The moral of the preceding chapter is that the main drawback of graphical models is their inability to describe all probabilistic conditional independence structures. This motivated an attempt to develop an alternative method for their description which overcomes that drawback and keeps some assets of graphical methods. The central notion of this method is the concept of a *structural imset* introduced in this chapter. Note that basic ideas of the theory were presented earlier [137] but (later recognized) superfluous details affected understanding of the message of the original series of papers. This monograph brings (in the next four chapters) a much simpler presentation, supplemented by new facts and perspectives.

4.1 Basic class of distributions

The class of probability measures for which this approach is applicable, that is, those whose induced conditional independence models can be described by structural imsets, is relatively wide. It is the class of *measures over N with finite multiinformation* mentioned in Section 2.3.4. The aim of this section is to show that this class involves three basic classes of measures used in practice in artificial intelligence and multivariate statistics.

4.1.1 Discrete measures

These simple probability measures (see Remark 2.2, p. 11) are mainly used in probabilistic reasoning [100] which is an area of artificial intelligence. Positive discrete probability measures are at the core of the models used in the analysis of contingency tables (see [70], Chapter 4), which is an area of statistics. The fact that every discrete probability measure over N has finite multiinformation is evident.

4.1.2 Regular Gaussian measures

These measures (see Section 2.3.6 for their basic properties) are widely used in mathematical statistics, in particular in multivariate statistics [28]. Corollary 2.6 says that every regular Gaussian measure over N has finite multiinformation.

4.1.3 Conditional Gaussian measures

This class of measures was proposed by Lauritzen and Wermuth [67] with the aim to unify discrete and continuous graphical models. In this book, their original class of measures is slightly extended. A *conditional Gaussian measure* P over N, called a *CG-measure over* N, will be specified as follows. The set N is partitioned into the set Δ of *discrete variables* and the set Γ of *continuous variables.* For every $i \in \Delta$, X_i is a finite non-empty set and $\mathcal{X}_i = \mathcal{P}(\mathsf{X}_i)$. For every $i \in \Gamma$, $\mathsf{X}_i = \mathbb{R}$ and $\mathcal{X}_i$ is the class of Borel sets in $\mathbb{R}$. A (discrete) probability measure P_Δ on $(\mathsf{X}_\Delta, \mathcal{X}_\Delta)$ is given. Moreover, provided that $\Gamma \neq \emptyset$, a vector $\boldsymbol{e}(x) \in \mathbb{R}^\Gamma$ and a positive definite $\Gamma \times \Gamma$-matrix $\boldsymbol{\Sigma}(x) \in \mathbb{R}^{\Gamma \times \Gamma}$ is ascribed to every $x \in \mathsf{X}_\Delta$ with $P_\Delta(x) > 0$. Then P is simply determined by its marginal for Δ and by the conditional probability on X_Γ given Δ:

$$P^\Delta \equiv P_\Delta,\;\; P_{\Gamma|\Delta}(\star\,|\,x) \equiv \mathcal{N}(\boldsymbol{e}(x), \boldsymbol{\Sigma}(x)) \quad \text{for every } x \in \mathsf{X}_\Delta \text{ with } P_\Delta(x) > 0\,.$$

Of course, these two requirements determine a unique probability measure on $(\mathsf{X}_N, \mathcal{X}_N)$. The above definition collapses in the case $\Gamma = \emptyset$ to a discrete measure over N and in the case $\Delta = \emptyset$ to a regular Gaussian measure over N.

Remark 4.1. Note that *positive CG-measures* (i.e., such that $P_\Delta(x) > 0$ for every $x \in \mathsf{X}_\Delta$) are mainly used in practice. A CG-measure of this type can be defined (see Lauritzen [70], § 6.1.1) by its density f with respect to the product of the counting measure on X_Δ and the Lebesgue measure on X_Γ

$$f(x, \boldsymbol{y}) = \exp{}^{\,\boldsymbol{g}(x) + \boldsymbol{h}(x)^\top \cdot \boldsymbol{y} - \frac{1}{2} \cdot \boldsymbol{y}^\top \cdot \boldsymbol{\Gamma}(x) \cdot \boldsymbol{y}} \qquad \text{for } \;x \in \mathsf{X}_\Delta,\, \boldsymbol{y} \in \mathsf{X}_\Gamma\,,$$

where $\boldsymbol{g}(x) \in \mathbb{R}$, $\boldsymbol{h}(x) \in \mathbb{R}^\Gamma$ and positive definite matrices $\boldsymbol{\Gamma}(x) \in \mathbb{R}^{\Gamma \times \Gamma}$ are named *canonical characteristics of* P. One can compute them directly from parameters $P_\Delta(x), \boldsymbol{e}(x), \boldsymbol{\Sigma}(x)$ which are named *moment characteristics* of the CG-measure as follows (see Lauritzen [70], p. 159):

$$\boldsymbol{\Gamma}(x) = \boldsymbol{\Sigma}(x)^{-1}, \qquad \boldsymbol{h}(x) = \boldsymbol{\Sigma}(x)^{-1} \cdot \boldsymbol{e}(x),$$

$$\boldsymbol{g}(x) = \ln P_\Delta(x) - \frac{|\Gamma|}{2} \cdot \ln(2\pi) - \frac{1}{2} \cdot \ln(\det(\boldsymbol{\Sigma}(x))) - \frac{1}{2} \cdot \boldsymbol{e}(x)^\top \cdot \boldsymbol{\Sigma}(x)^{-1} \cdot \boldsymbol{e}(x).$$

These measures are positive in the sense of Section 2.3.5 but they do not involve all discrete measures. For this reason, the original class of CG-measures from Lauritzen and Wermuth [67] has been slightly extended in this book. $\triangle$