Graphical Models for Multivariate Time Series

Giulia Pellegrino
Registration Number: 612145
Supervisor: Professor Michele Pavon
University of Padua
Faculty of Engineering
Bachelor Degree in Information Engineering

Contents

Chapter 1

Introduction

The graph theoretic side of graphical models provides both an intuitively appealing interface by which humans can model highly-interacting sets of variables as well as a data structure that lends itself naturally to the design of efficient general-purpose algorithms.

— Michael Jordan, 1998.

The aim of this thesis is to give a brief introduction to the *Graphical Models for Multivariate Time Series* and show some of its applications in order to emphasize the importance of these models.
In the second chapter, we recall some notation relevant to the introduction of *Graphical Models* such as the concept of *Inverse Covariance Matrix* and *Multivariate Power Spectrum*.
Fields such as bioinformatics, information retrieval, speech processing and communications often involve large-scale models in which thousand or millions of random variables are linked in complex ways. Graphical models provide a general methodology to approach these problems.
Probabilistic graphical models use a graph-based representation as the basis for compactly encoding a complex distribution over a high-dimensional space. In this graphical representation the nodes correspond to the variables in our domain, and the edges correspond to direct probabilistic interaction between them.
In Chapter 3, we present Graphical Models in the static case, in which the characterization of all conditional independence relations is given by a specific pattern of zeros in the inverse covariance Σ^{-1}.
In Chapter 4, we extend this concept to multivariate time series. Using the partial spectral coherence between components given the rest of the components, we understand that if there is the lack of a connecting link between two nodes it signifies conditional independence between two processes given the remaining.

In other words, graphical models give a graph representation of relations between random variables and processes. The simple example is a Gaussian graphical model in which an undirected graph with n nodes is used to described conditional independence relations between the components of an n-dimensional random variable $x \sim \mathcal{N}(0,\ \Sigma)$.
The absence of an edge between two nodes of the graph indicates that the corresponding components of x are conditional independent given the others, that corresponds to zero entries in the inverse covariance matrix.
Conditional independence between components of a multivariate stationary Gaussian process can be characterized in terms of the inverse of the spectral density matrix $\mathbf{\Phi}(e^{i\theta})$, in which two components $x_k(t)$ and $x_l(t)$ are independent of the other components if and only if

$$[\mathbf{\Phi}(e^{i\theta})^{-1}]_{kl} = 0.$$

The problem of estimating the topology in a Gaussian graphical model is more involved. An approach is to formulate hypothesis testing problems to decide about the presence or absence of edges between two nodes. Another possibility is to enumerate different topologies, and use information criteria to rank the models.
In Chapter 5, we address the extension of estimation methods for Gaussian graphical models to autoregressive (AR) processes. Conditional independence constraints in AR estimation methods are enforced by placing restrictions on the sparsity pattern of the inverse spectral density matrix.
We show that AR models can be approximated by a low order ARMA model and the benefits of this approximation.
In the final Chapter, we present an example of a biomedical application in intensive care because graphical models provide information on the relationship among physiological variables that are helpful in many medical problems.

Chapter 2

Notation and Definition

In this chapter, we introduce some definitions in preparation for the concept of *Graphical Models* for Gaussian (normal) random variables and for multivariate time series.

2.1 Probability Theory

Definition 2.1 (Random Variable)**.** A probability space consists of three parts: a sample space, Ω, which is the set of all possible outcomes; a set of collection of all such events is a σ-algebra $\mathcal{F}$; the assignment of probabilities to the events, that is, a function from events to probability levels $\mathbb{P}$.
Suppose that $(\Omega,\ \mathcal{F},\ \mathbb{P})$ is a probability space. A function

$$\mathbf{X} : \Omega \to \mathbb{R},$$

is called a *random variable*, if for all $x \in \mathbb{R}$, $\{\omega \in \Omega : \mathbf{X}(\omega) \leq x\} \in \mathcal{F}$.
Random variables can be classified as *discrete*, namely variables that have finitely or countably many values, or *continuous*, namely variables that can take any value within a continuous range.

Example 2.1.1. Consider an experiment where a coin is tossed three times. If X represents the number of times that the coin comes up heads, then X is a discrete random variable that can only have the values 0,1,2,3.

An example of a continuous random variable would be an experiment that involves measuring the amount of rainfall in a city over a year, or the average height of a group of 25 people.

Definition 2.2 (Probability Distribution)**.** Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\mathbf{X}$ is a random variable $\mathbf{X} : \Omega \to \mathbb{R}$. That the *probability distribution function* for $\mathbf{X}$ is defined by $F_{\mathbf{X}}(x) = P(\mathbf{X} \leq x)$.

Definition 2.3 (Probability Density)**.** For a random variable taking values in $\mathbb{R}$, if $F_{\mathbf{X}}(x)$ is differentiable, then the derivative with respect to x of $F_{\mathbf{X}}(x)$, denoted with a lowercase letter f

$$f_x(x) = \frac{d}{dx} F_{\mathbf{X}}(x)$$

is called the *probability density* function or simply the *probability density* of **X**. It satisfies the conditions:

$$f(x) \geq 0,$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

A random variable has an associated *expected value*, *variance* and *covariance*.

Definition 2.4 (Expected Value)**.** The *Expected value* of a random variable indicates its average or central value. It is a useful summary value (a number) of the variable distribution. The expected value of a random variable X is symbolized by E(X) or μ.

If X is a discrete random variable with possible values x_1, x_2, x_3, ..., x_n, ... and P(X = x_i) denotes "the probability that the outcome x_i occurs", then the expected value of X is defined by:

$$E(x) = \mu = \sum_{i=1}^{\infty} x_i P\{X = x_i\} \tag{2.1}$$

provided the series converges.

If X is a continuous random variable with probability density function f(x), then the expected value of X is defined by:

$$E(x) = \mu = \int_{-\infty}^{\infty} x f(x)dx \tag{2.2}$$

Example 2.1.2. Discrete case : When a die is rolled, each of the possible faces 1, 2, 3, 4, 5, 6 (the x_i's) has a probability of 1/6 (the p(x_i)'s) of showing. The expected value of the face showing is therefore:
μ = E(X) = (1 x 1/6) + (2 x 1/6) + (3 x 1/6) + (4 x 1/6) + (5 x 1/6) + (6 x 1/6) = 3.5
Notice that, in this case, E(X) is not a possible value of X.

Definition 2.5 (Variance and Covariance)**.** The *Variance* of a random variable is a non-negative number that gives an idea of how spread around the

mean the values of the random variable are likely to be; the larger the variance, the more scattered the observations on average. It is symbolized by V(X) or Var(X) or σ^2.
The variance of the random variable X is defined to be:

$$Var(X) = \sigma^2 = E(X - E(X))^2 = E(X^2) - E(X)^2 \tag{2.3}$$

where E(X) is the expected value of the random variable X.

If X and Y are two random variables with $E(X^2) < \infty$ and $E(Y^2) < \infty$, then their *Covariance* is defined by:

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y). \tag{2.4}$$

Definition 2.6 (Random Vector)**.** We say that X = (X_1, ... ,X_n) is an n-dimensional random vector if its components are random variables. This is equivalent to saying that X is a random variable with values in $\mathbb{R}^n$.

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \in \mathbb{R}^n \qquad \mathbf{X} : \Omega \to \mathbb{X} \tag{2.5}$$

The mathematical expectation of an n-dimensional random vector X is, by definition, the vector

$$E(X) = (E(X_1), ..., E(X_n))^T$$

The *Covariance Matrix* (or *Variance Matrix*) of an n-dimensional random vector X, denoted by Σ, is defined as follows:

$$\Sigma_{ij} = Cov(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] \tag{2.6}$$

where $\mu_i = \mathrm{E}(X_i)$ is the expected value of the ith entry in the vector X.

In other words, we have

$$\Sigma = \begin{bmatrix} E[(X_1 - \mu_1)(X_1 - \mu_1)] & E[(X_1 - \mu_1)(X_2 - \mu_2)] & \cdots & E[(X_1 - \mu_1)(X_n - \mu_n)] \\ E[(X_2 - \mu_2)(X_1 - \mu_1)] & E[(X_2 - \mu_2)(X_2 - \mu_2)] & \cdots & E[(X_2 - \mu_2)(X_n - \mu_n)] \\ \vdots & \vdots & \ddots & \vdots \\ E[(X_n - \mu_n)(X_1 - \mu_1)] & E[(X_n - \mu_n)(X_2 - \mu_2)] & \cdots & E[(X_i - \mu_i)(X_j - \mu_j)] \end{bmatrix}$$

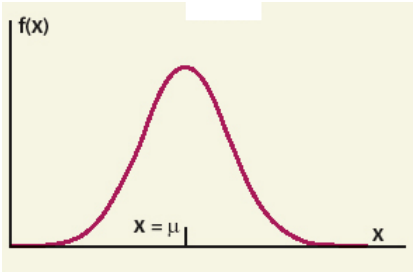
The inverse of this matrix, Σ^{-1}, is the *Inverse Covariance Matrix*.

Definition 2.7 (Discrete Time Stochastic Process and Multivariate Stochastic Process)**.** In statistics, signal processing, econometrics and mathematical finance, a *Time Series or Discrete Time Stochastic Process* is a sequence of data points, measured at successive time instants spaced at uniform time intervals.
Let $(\Omega,\ \mathcal{F},\ \mathbb{P})$ be a probability space. A family of random variables $\{X(t); t \in T\}$ defined on Ω is called a *Stochastic Process.* It is called a *continuous-time* process if $T = [a, b], -\infty \leq a < b \leq +\infty$. It is called a *discrete-time* process if $T \subseteq \mathbb{Z}$.
In the simple case of discrete time, a stochastic process amounts to a sequence of random variables known as *Time Series.*
Thus, such a process is the family of random variables X_1, X_2 ,... and to know it, it is necessary to specify the functions of joint probability density of every subset of them. Formally a process $\{X_t\}$ is known if the joint distribution of $(X_{t_1}\ ,\ X_{t_2}\ ,\ \ldots\ ,\ X_{t_k})$ is known for every k and for every k-tuple of values $(t_1\ ,\ t_2\ ,\ \ldots\ ,t_k)$.

A *Multivariate Stochastic Process* is a n-dimensional *stochastic process* $\{X_t\}$, such that X_t is a random vector $\forall t$.

Definition 2.8 (Gaussian Distribution)**.** The *Normal* or *Gaussian* distribution is a continuous probability distribution that has a bell-shaped probability density function, known as the Gaussian function or informally the bell curve:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2.7}$$



where the parameter μ is the mean and σ^2 is the variance. The distribution with $\mu = 0$ and $\sigma^2 = 1$ is called the standard normal distribution or the unit normal distribution. The Normal distribution is important also because it can be completely specified by two parameters: mean and variance.
This is the reason why we denote a variable with Normal distribution as $\mathbf{X} \sim \mathcal{N}(\mu,\ \sigma^2)$.

While the univariate Gaussian is defined in terms of two parameters, a mean and a variance, a multivariate Gaussian distribution of $X_1, ..., X_n$ is

characterized by an n-dimensional *mean vector* $\boldsymbol{\mu}$, and a symmetric $n \times n$ *covariance matrix* Σ, which we assume to be nonsingular. The density function is defined as:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right] \tag{2.8}$$

where $|\Sigma|$ is the determinant of Σ.
We can observe the presence of the inverse of covariance matrix Σ^{-1} in this formula.

Definition 2.9 (Autocorrelation Function)**.** The *Autocorrelation Function* of a random process $\mathbf{X}$ is a function of two real variables, the reference time t and the delay time (lag) τ. It takes the value of the correlation between two random variables in the process, one taken at instant t, and the other at the instant t-τ.

$$\mathbf{r}_x(t, t-\tau) = E[x(t)x(t-\tau)^T] \tag{2.9}$$

A process is called *wide-sense stationary* if its mean, variance and autocorrelation function are independent of time, depending only on τ that is the time lag.

$$\mathbf{r}_x(t, t-\tau) = \mathbf{r}_x(t-(t-\tau), 0) = \mathbf{r}_x(\tau, 0)$$

Then the important characteristic of autocorrelation function is that a time shift of a signal does not change its autocorrelation.

Definition 2.10 (Power Spectrum)**.** The power spectrum is a matrix-valued function in the frequency domain and it is defined as the Fourier transform of the auto-correlation function:

$$\Phi(e^{i\theta}) = \sum_{k=-\infty}^{+\infty} R_k e^{-ik\theta} \tag{2.10}$$

where $R_k = E\{x(k)x(0)^T\}$ are named "Covariance lags".

Since R_k has the Hermitian Property:
$R_{-k}^T = E\{x(k-k)x(0-k)^T\}^T = E\{x(0)x(k)^T\}^T = E\{x(k)x(0)^T\} = R_k$,
we have:

$$\Phi(e^{-i\theta})^T = \sum_{k=-\infty}^{+\infty} R_k^T e^{ik\theta} = \sum_{h=-\infty}^{+\infty} R_{-h}^T e^{-ih\theta} = \sum_{h=-\infty}^{+\infty} R_h e^{-ih\theta} = \Phi(e^{i\theta}).$$

Namely, Φ enjoys the *Parahermitian property.*

Definition 2.11 (Independent Random Variables)**.** Formally, the *conditional probability* of an event B given A is defined as

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

if P(A) > 0.

Two random variables X and Y, are said to be independent if the occurring of the value x of X has no influence on the occurring of the values of Y, and vice versa.
For discrete independent random variables, their probabilities are related by:

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$$

for each pair (x_i, y_j).

Anyway, two events A and B are *conditionally independent* given a third event C precisely if and only if, given knowledge of whether C occurs, knowledge of whether A occurs provides no information on the likelihood of B occurring, and knowledge of whether B provides no information on the likelihood of A occurring.
In the standard notation of probability theory, A and B are conditionally independent given C if and only if

$$P(A \cap B|C) = P(A|C)P(B|C).$$

2.2 Graphs

In this section, we survey some of the basic concepts in graph theory used in the following chapters.

Definition 2.12 (Nodes and Edges)**.** A graph is a data structure $\mathcal{K}$ consisting of a set of nodes and edges. In this thesis, we will assume that the set of nodes is $\mathcal{X} = \{X_1, ..., X_n\}$. An edge in a graph is a connection between vertices, a set of edges is usually denoted E.

Definition 2.13 (Directed and Undirected)**.** A pair of nodes X_i, X_j can be connected by a *directed edge* if the edge have a direction,$X_i \rightarrow X_j$, that means an edge starting at X_i and going to X_j or an *undirected edge* if the edges have no implied direction $X_i - X_j$, that is the same as $X_j - X_i$.

Chapter 3

Graphical Models

Graphical models expose the conditional independences between families of probability distributions in the form of graphs. We now consider that Gaussian (normal) random variables involve the specification of a structure on the mean and the covariance matrix Σ.
We put our interest on the structure of Σ^{-1} in which certain elements are assumed to be zero and we show how these zeros in Σ^{-1} correspond to a conditional independence relation of two random variables given the rest of the components. Such process can be represented on a graph where the random variables are nodes and the lack of a connecting link between two nodes signifies *conditional independence*.

3.1 Conditional Independence

We consider a random vector $\mathbf{X}$ having a Gaussian distribution with mean 0 and positive definite covariance matrix $\mathbf{\Sigma}$. The components of $\mathbf{X}$ will be indexed by a finite set C and for $a \subset C$, we write $\mathbf{X}_a$ for the subset of the components of $\mathbf{X}$ indexed by a, namely $(\mathbf{X}_\gamma \colon \gamma \in a)$. The covariance matrix $\Sigma = (\Sigma(\alpha, \beta) : \alpha, \beta \in C)$ on C is defined by $\Sigma(\alpha, \beta) = \mathrm{E}\{X_\alpha X_\beta\}$, where E denoted expected value and $\alpha, \beta \in C$.
For subset $a, b \subseteq C, \Sigma_{a,b} = \{\Sigma(\alpha, \beta) : \alpha \in a, \beta \in b\}$ denotes the cross covariance matrix of $\mathbf{X}_a$ and $\mathbf{X}_b$. When $a = b$ we write directly Σ_a instead of $\Sigma_{a,a}$. We abbreviate the set intersection $a \cap b$ to ab and write $a \setminus b$ for the complement of b in a.
The following proposition relates the conditional independence of two components of $\mathbf{X}$ to the structure of Σ.

Proposition 3.1.1. *For subsets a, b of C with $a \cup b = C$ the following statements are equivalent:*

(i) $\Sigma_{a,b} = \Sigma_{a,\ ab}\Sigma_{ab}^{-1}\Sigma_{ab,\ b}$

(ii) $\Sigma_{a\setminus b,b\setminus a} = \Sigma_{a\setminus b,\ ab}\Sigma_{ab}^{-1}\Sigma_{ab,\ b\setminus a}$

(iii) $(\Sigma^{-1})_{a\setminus b,\ b\setminus a} = 0$

(iv) $\mathbf{X}_a$ *and* $\mathbf{X}_b$ *are conditionally independent given* $\mathbf{X}_{ab}$

Proof. We can easily see that if two random variables are independent their covariance is null. In the same way, we set the standard formula 3.1 for the conditional covariance matrix null and obtain the connection between (iv) and (ii).
(i) and (ii) are equivalent by partioning the rows of Σ over $a \setminus b$ and ab and the columns over $b \setminus a$ and ab.
By partitioning over $a\setminus b$, $b\setminus a$ and ab, a straightforward use of the expression for the inverse of a partitioned matrix proves that (ii) is equivalent to (iii).

$$cov(\mathbf{X}_{a\setminus b}, \mathbf{X}_{b\setminus a}|\mathbf{X}_{ab}) = \Sigma_{a\setminus b,\ b\setminus a} - \Sigma_{a\setminus b,\ ab}\Sigma_{ab}^{-1}\Sigma_{ab,\ b\setminus a} \tag{3.1}$$

□

More precisely we are interested in the next corollary that follows from 3.1.1 put $a = C \setminus \{\alpha\}' and b = \{\beta\}'$, where the set C \ b will be denoted b':

Corollary 3.1.2. *For distinct elements α, β of C, $\mathbf{X}_\alpha$ and $\mathbf{X}_\beta$ are conditionally independent given $\mathbf{X}_{\{\alpha,\beta\}'}$ if and only if $\Sigma^{-1}(\alpha, \beta) = 0$*

Now, after we have shown that zeros in Σ^{-1} correspond to conditional independence relations, we associate an undirected graph to the pattern of zeros.

3.2 Undirected Graph

An undirected graph will be denoted by $\mathcal{G} = (V, E(\mathcal{G}))$ where V is the vertex set and $E(\mathcal{G})$ the edge set. For any vertex γ we write $\partial\gamma = \{\alpha : \{\alpha, \gamma\} \in E(\mathcal{G})\}$ for the set of neighbours of γ.
The independence model for undirected or bi-directed graphs is described using the *global Markov property* defined as follows. Let $\mathcal{G}$ be a undirected graph with vertex set V. We say that set V_i is independent of V_j given V_k if V_i is separed from V_j by $V\setminus(V_i \cup V_j \cup V_k)$.
We denote the independence implied by the global Markov property on a

undirected graph $\mathcal{G}$ by $V_i \perp_{\mathcal{G}} V_j | V_k$. The characterization of all conditional independence relations following from a given pattern of zeros in Σ^{-1} is presented in 3.2.1.

Proposition 3.2.1. *Let $\mathcal{G}$ be a simple graph with vertex set V indexing the Gaussian random variables $\mathbf{X}$. Then the following are equivalent:*

(i) $\Sigma^{-1}(\alpha, \beta) = 0$ if $\{\alpha, \beta\} \notin E(\mathcal{G})$ and $\alpha \neq \beta$;

The local Markov property:

(ii) For every $\gamma \in V$, X_γ and $\mathbf{X}_\gamma$, are conditionally independent given $\mathbf{X}_{\partial\gamma}$;

The global Markov property:

(iii) For every a, b and d with d separating a from b in $\mathcal{G}$, $\mathbf{X}_a$ and $\mathbf{X}_b$ are conditionally independent given $\mathbf{X}_d$.

Proof. (i) is equivalent to (ii) put $a = \{\bar{\gamma}\}$ and $b = \{\gamma\}'$ and $\Sigma^{-1}(\gamma, \{\bar{\gamma}\}') = 0$ proves the result. The equivalence between (i) and (iii) is proved put "a" $= a \cup d$ and "b" $= b \cup d$ with $a \subseteq a^*$, $b \subseteq b^*$, $a^* \cup b^* \cup d = V$ and a^* is separated from b^* by d, now the integration to obtain the marginal density of $\mathbf{X}_{a \cup b \cup d}$ shows that (i) implies (iii).
At last if $(\alpha, \beta) \notin E(\mathcal{G})$ then $\alpha,\ \beta$ are separated by $\{\alpha,\ \beta\}'$, hence by (iii) X_α and X_β are conditionally independent given $X_{\{\alpha,\beta\}'}$ and Corollary 3.1.2 shows that $\Sigma^{-1}(\alpha, \beta) = 0$. □

Example 3.2.1. Suppose Σ^{-1} has the following pattern with $\star$ denoting a nonzero element:

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} \star & \star & 0 & 0 & 0 \\ \star & \star & \star & 0 & \star \\ 0 & \star & \star & \star & \star \\ 0 & 0 & \star & \star & \star \\ 0 & \star & \star & \star & \star \end{bmatrix} \end{array}$$

The corresponding graph $\mathcal{G}$ would be as shown in the Figure 3.1. If we put $\gamma = \{2\}$, $\partial\gamma = \{1, 3, 5\}$, and use (ii) we deduce that X_2 and X_4 are conditionally independent given $\mathbf{X}_{\{1,3,5\}}$. In the same way with $a = \{1\}$, $b = \{4\}$ and $d = \{2\}$, the property (iii) asserts that X_1 and X_4 are conditionally independent given X_2.

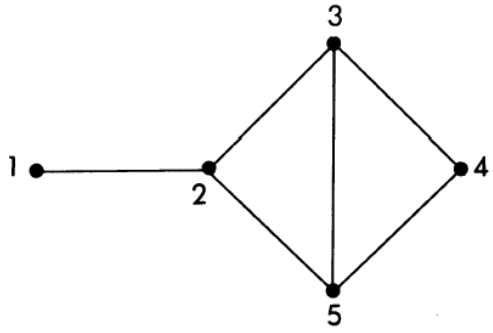


Figure 3.1: Example

3.3 Static Application

Consider a very simple medical diagnosis setting, where we focus on two diseases, flu and hayfever; these are not mutually exclusive, as a patient can have either, both, or none. Thus, we might have two binary-valued random variables, Flu and Hayfever. We also have a 4-valued random variable Season, which is correlated both with Flu and Hayfever. We may also have two symptoms, Congestion and Muscle Pain, each of which is also binary-valued. Overall, our probability space has $2 \quad 2 \quad 4 \quad 2 \quad 2 = 64$ values, corresponding to the possible assignments to these five variables.
When we consider the fact that a typical medical-diagnosis problem has dozens or even hundreds of relevant attributes, the problem appears completely intractable. Graphical models provide a mechanism for exploiting structure in complex distributions to describe this problems compactly and allow them to be constructed and utilized effectively.

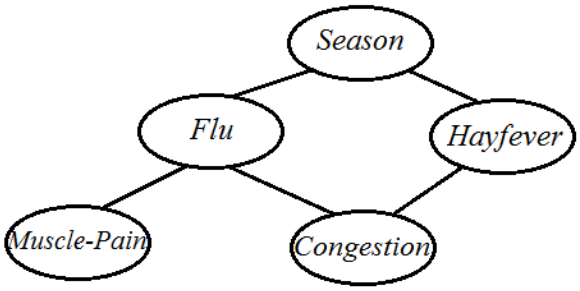


Figure 3.2: Probabilistic graphical model of Flu

Independencies

$$(F \perp H \mid S)$$

$$(C \perp S \mid F, H)$$

$$(M \perp H, C \mid F)$$

$$(M \perp C \mid F)$$

$$\begin{array}{c} \\ S \\ F \\ H \\ M \\ C \end{array} \begin{array}{c} \begin{array}{ccccc} S & F & H & M & C \end{array} \\ \begin{bmatrix} \star & \star & \star & 0 & 0 \\ \star & \star & 0 & \star & \star \\ \star & 0 & \star & 0 & \star \\ 0 & \star & 0 & \star & 0 \\ 0 & \star & \star & 0 & \star \end{bmatrix} \end{array}$$

In this graph, we see that there is no direct interaction between *Muscle Pain* and *Season*, but both interact directly with *Flu*.
The graph is a compact represantation of a set of independencies that hold in the distribution.
The distribution encoding our beliefs about this particular situation may satisfy the conditional independence (*Congestion* $\perp$ *Season* | *Flu, Hayfever*). Note that this assertion implies only that all of the information we may obtain from the season on the chances of having congestion we already obtain by knowing whether the patient has the flu and has hayfever.

Chapter 4

Graphical Models for Multivariate Time Series

Now we extend the concept of *Graphical Models* for multivariate data to multivariate time series. We use the partial spectral coherence between two components given the remaining components to define a partial correlation graph for time series.
The edges of a conditional independence graph show the conditional dependence structure between several variables and give an idea of the interaction structure of the observed variables.
In regard of the concept of graphical models for multivariate time series, the vertex set will consist of the *components of the series* while the edge will reflect the partial correlation structure of the components given the others. In order to identify the graph we show a property of the inverse of the spectral matrix.

4.1 Graphical models for time series

We define a graph $G = (V, E)$ that consists of a set of vertices V, say $V = \{1, ..., m\}$ and a set of edges $E \subset \{(a, b) \in V \times V\}$. We only consider undirected graphs, i.e. we assume $(a, b) \in E$ whenever $(b, a) \in E$.
The core idea is that an edge (a, b) is missing if the components $X_a(\cdot)$ and $X_b(\cdot)$ are uncorrelated given the other components of the series, in the same way as in the static case.
Now consider an m-dimensional, zero-mean, Gaussian, stationary stochastic process $\{x(t)\}_{t\in\mathbb{Z}}$ with the property that designated pairs of components are conditionally independent given the rest of the components. In fact, such processes can be represented on a graph where the components are nodes

and the lack of a connecting link between two nodes signifies conditional independence.
This is manifested by a sparsity pattern in the inverse of the $m \times m$ matrix-valued spectral density

$$\mathbf{\Phi}(e^{i\theta}) = \sum_{k=-\infty}^{+\infty} R_k e^{-ik\theta} \tag{4.1}$$

where

$$R_k = E\{x(k)x(0)^T\}$$

and where we assume that $\mathbf{\Phi}(e^{i\theta}) > 0$ for all $\theta \in [-\pi, \pi]$.
As it is demonstrated in [1], we have

$$[\mathbf{\Phi}(e^{i\theta})^{-1}]_{kl} = 0, \qquad -\pi \leq \theta \leq \pi \tag{4.2}$$

for pairs (k, l) such that x_k and x_j are conditionally independent give the rest of the components of the process x.
More precisely, given $V = \{1, 2, ..., m\}$ and

$$X_I = span\{x_j(t) : j \in I, t \in \mathbb{Z}\}$$

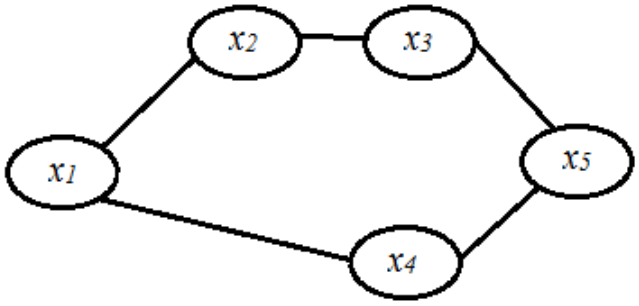
for an arbitrary set $I \subset V$, the (4.2) holds for all pairs (k, l) such that $X_{\{k\}}$ and $X_{\{l\}}$ are *conditionally independent* given $X_{V\setminus\{k,l\}}$, which we write as

$$X_{\{k\}} \perp X_{\{l\}} | X_{V\setminus\{k,l\}}.$$

The set of all such conditional independence relation constitutes a graph $G = (V, E)$ where V, defined as before, is the set of vertices and $E \subseteq V \times V$ is a set of edge defined in the following way

$$(k, l) \notin E \iff k \neq l, \quad X_{\{k\}} \perp X_{\{l\}} | X_{V\setminus\{k,l\}}.$$

An example is depicted in the following Figure:



where the lack of an arc between nodes k and l signifies conditional independence between the process $\{x_k(t)\}_{\mathbb{Z}}$ and $\{x_l(t)\}_{\mathbb{Z}}$ given the rest of the components processes.

A model of the process x which takes conditional independence relations into consideration is commonly referred to as a *graphical model*.

Chapter 5

Arma and Graphical Models

5.1 Models

Models for time series data can have many forms and represent different stochastic processes. When modeling variations at the level of a process, two broad classes of practical importance are the *autoregressive* (AR) models and the *moving average* (MA) models. Combinations of the two produce *autoregressive moving average* (ARMA).

Definition 5.1 (Autoregressive Model)**.** the AR(p) model is written

$$X(t) = c \ + \ \sum_{j=1}^{p} A_j X(t-j) - e(t) \tag{5.1}$$

where $A_1, ..., A_p$ are the *autoregression coefficients* of the model, c is a constant and $e(t)$ is white noise. The constant term is omitted by many authors for simplicity.
An autoregressive model is essentially a model which depends only on the previous outputs of the system $X(t-1), X(t-2), ...$ In other words, the current term of the series can be estimate by a linear weighted sum of previous terms in the series. The weights are the autoregression coefficients.
By convention the series X_t is assumed to be zero mean, if not this is simply another term A_0 in front of the summation in the equation above.

Definition 5.2 (Moving-Average Model)**.** The notation MA(q) refers to the moving average model of order q:

$$X(t) = \mu + e(t) + \sum_{j=1}^{q} B_j e(t-j) \tag{5.2}$$

where $B_1, ..., B_q$ are the parameters of the model, μ is the expectation of $X(t)$ (often assumed to equal 0), and the $e(t-1), e(t-2), ...$ are again, white

noise error terms.

Definition 5.3 (ARMA Models)**.** Combining AR and MA models, we can define *Autoregressive Moving-Average Models* or ARMA(p,q) models as:

$$X(t) = c + e(t) + \sum_{j=1}^{p} A_j X(t-j) + \sum_{j=1}^{q} B_j e(t-j) \tag{5.3}$$

with p the order of the AR part, and q the order of the MA part.

5.2 ARMA Identification of Graphical Models

Gaussian graphical models represent a way to handle the conditional independence relations of a set of stochastic variables or processes. Such models induce a sparsity pattern in the spectrum inverse.
In order to apply this result to AR processes, we need to express the inverse spectrum in terms of the model parameters, but the problem of estimating the model parameters A from an observed sequence $X(t-1), ..., X(t-p)$ is not always simple.
There are problems to fit an autoregressive (AR) model to such a process as a mean for assessing conditional independence.
The basic idea is to use a maximum likelihood and ask for consistency of the AR model with the data together with conditional independence between particular nodes. In particular, given the (estimates of) autocovariances $R_0, R_1, ..., R_n$, the problem is to find a multivariate autoregressive model

$$\sum_{j=0}^{n} A_j X(t-j) = e(t)$$

that satisfies the sparsity condition seen in Chapter 4:

$$[\mathbf{\Phi}(e^{i\theta})^{-1}]_{kl} = 0, \qquad -\pi \leq \theta \leq \pi$$

Here $\{e(t)\}_{t\in\mathbb{Z}}$ is a white noise process and $A_0, A_1, ..., A_n$ are $m \times m$ matrices such that the determinant of the matrix polynomial have no zeros in the closed unit disc and $\det A_0 > 0$.
In any case, there are examples where AR models are insufficient. Furthermore, an AR model of exceedingly high order can be approximated by a low order ARMA model. Now we bring this problem to an advanced level, that is to fit an autoregressive moving-average (ARMA) model, while respecting the sparsity constraint, to the same data. We follow [3, 69-72].
By allowing for ARMA models, we may choose from a continuum of infinitely many solutions, one of which might satisfy the required graph topology better.

The ARMA models that we shall consider here take the form

$$\sum_{j=0}^{n} A_j X(t-j) = \sum_{j=0}^{n} B_j e(t-j). \tag{5.4}$$

For technical motivations, we shall here assume that the matrix coefficients of the moving-average part has the form

$$B_j = b_j I, \qquad j = 0, 1, ..., n, \qquad b_0 = 1.$$

where the scalar polynomial

$$b(z) = z^n + b_1 z^{n-1} + ... + b_n$$

has no zeros in the closed unit disc.
Then the spectral density of the stationary vector process X becomes $\Phi(z) = W(z)W(z^{-1})^T$, where $W(z) = b(z)A(z)^{-1}$.
Hence, our basic problem is to determine a spectral density of the form

$$\Phi(z) = \psi(z)Q(z)^{-1} \tag{5.5}$$

satisfying the sparsity constraint and the moment conditions

$$\int_{-\pi}^{\pi} e^{ik\theta}\Phi(e^{i\theta})\frac{d\theta}{2\pi} = R_k, \qquad k = 0, 1, ..., n$$

where ψ is a scalar pseudo-polynomial of degree at most n and Q is a symmetric $m \times m$ matrix-valued pseudo-polynomial of degree n. Then the coefficients in the corresponding ARMA model 5.3 can be obtained by determining the minimum-phase spectral factors A(z) and b(z) from

$$A(z)A(z^{-1})^T = Q(z) \qquad and \qquad b(z)b(z^{-1}) = \psi(z) \tag{5.6}$$

respectively.

In applied problems, we are just given a string of measured data

$$x_0, x_1, ..., x_N \in \mathbb{R}^n$$

from the ARMA model 5.3, and we want to estimate the parameters A_0, A_2, $..., A_n, b_1, ..., b_n$ without prior knowledge of the topology of the graph. Hence we also need to estimate a suitable graphical structure E from the data. In fact, in many applications, determining the topology of the graph is the main task.

For some applications, the interaction graphs have been determined by nonparametric methods. The ARMA approach also produce a dynamical model which can be used for prediction.

In other word, correlation data provide moment constraints on the matricial power spectrum of a graphical model. When the topology of the graph has been estimated [1], the conditional independence features impose certain zeros in the inverse of the spectral density. Present work suggests the use of autoregressive models for assessing conditional independence. It is possible using a maximum likelihood, or a maximum entropy formalism, and ask for consistency of the AR model with the data together with conditional independence between certain nodes.
However, since AR models of exceedingly high order can be approximated by a low order ARMA model, the potential benefit of analyzing the finite-dimensional manifold of solutions to moment problems instead is evident, as there are examples where AR models are insufficient. There is the possibility of *fitting a low order ARMA model* starting from the graph topology and correlation data.

Chapter 6

Biomedical applications

In intensive care, detection of critical states and of intervention effects is of great importance for suitable bedside decision support. Intelligent alarm systems are needed to provide suitable bedside decision support. Graphical models provide information on the relationships among physiological variables that are helpful, [6], [8, 340-344].
Distinct clinical states are characterized by distinct partial correlation structures. Hence, this technique can provide new insights into physiological mechanisms.

Clinical information systems can acquire and store physiological variables and device parameters online at least every minute.
In view of the high dimension of the data, severe problems arise from the natural limitations of human beings because it is difficult to develop a systematic response to problems involving more than seven variables, and human beings are not able to judge the degree of relatedness between more than two variables. Thus, besides the aim of detecting clinical states, reducing the number of variables is an added task.
We need information on the relationships between the variables. Graphical interaction models are an important tool for investigating and modeling relationships within multivariate data, as seen in the previous chapters. These models allow a simple and helpful graphical visualization, where the variables are represented by vertices and the relationships between the variables are illustrated by edges.

6.1 Data set

In the analysis of intensive care physiological variables of critically ill, we concentrate on the variables heart rate (HR), arterial diastolic pressure (APD), arterial systolic pressure (APS), arterial mean pressure (APM), pulmonary artery diastolic pressure (PAPD), pulmonary artery systolic pres-

sure (PAPS), pulmonary artery mean pressure (PAPM), central venous pressure (CVP), blood temperature (Temp) and pulsoximetry (SpO2).
These variables are important for the detection of critical, possibly life-threatening situations and they provide information on the clinical status of the patient.

6.2 Graphical interaction models

Between multiple variables usually a multitude of relationships exist, but many of them are indirect, i.e. they are induced by others. Statistical analysis in form of graphical models helps to reveal the essential relationships with no information loss.
We compare "empirical relationships" found by statistical analysis to "physiological relationships" based on medical knowledge.
Physiological relationships mean that a change in one physiological variable leads to a change in another physiological variable, e.g. an increase of systolic blood pressure close to always leads to an increase in mean blood pressure.
From a statistical point of view, measurements of physiological variables observed in short time intervals form multivariate time series as there may be interactions not only between the measurements observed at the same time point, but at short time lags, too. We use partial correlation graphs for multivariate time series, where linear relationships between every pair of variables at all time lags are investigated controlling for the linear effects of the other variables at all time lags. This allows to detect relationships between the variables of a multivariate time series.

6.3 Results

A first step for online monitoring is to find representative variables, in order to get reliable and interpretable results without substantial loss of information we need to understand the relationship between the variables. For this task, we analyze the relationships between all vital signs mentioned above. Figure 6.1 shows a typical example of a partial correlation graph for the hemodynamic system resulting from the analysis of the data measured for one patient, where different line types depict different strength of partial correlation.

For all patients strong relationships could be identified between the arterial pressures (APS, APD and APM), between the pulmonary artery pressures (PAPS, PAPD, PAPM) as well as between heart rate and pulse. Hence, we can identify groups of strongly related variables from an analysis of the

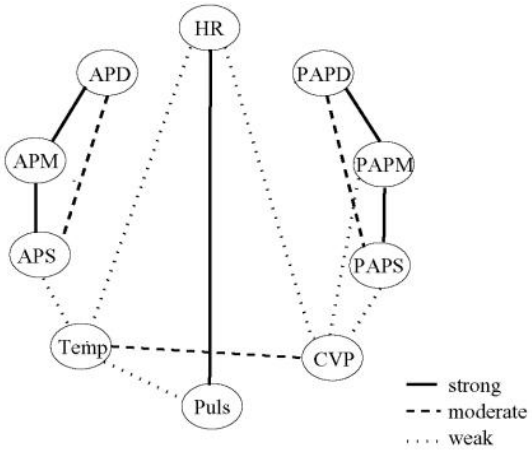


Figure 6.1: Partial correlation graph, one step selection. Different line types depict different strength of partial correlation.

full data sets. Further relationships could be identified for some patients, e.g. between arterial pressures and heart rate, and between pulmonary artery pressures and central venous pressure.

A partitioning of the variables into strongly related subgroups as given above can be used to reduce the number of variables which have to be considered for online monitoring. The absence of a relation between two variables VI and V2 means that the observations of V2 do not add anything to explain the course of variable VI (and vice versa) given the measurements of the remaining variables. On the other hand, if a variable has strong relationships to several other variables it provides a lot of information on these variables.

In the previous analysis we inspected the relationships using all variables. This may hide some relationships when there are groups of variables which are only slightly different representations of the same physiological process. When analyzing whether there is a relationship between PAPM and APM, we subtract the linear influences of all other variables including APD, APS, PAPD and PAPS. As a consequence, systolic and diastolic pressures are dropped in the following. In this way, a set of 'important variables' consisting of HR, APM, PAPM, CVP, SP02 and Temp is retained.

Distinct clinical states such as pulmonary hypertension, congestive heart failure and vasopressor support are accompanied by different pathophysiological responses of the circulatory system and we can see in Figure 6.2 that correspond to dictinct graphical models. These changes may be supposed to result in differences in the interactions between the vital signs, too.

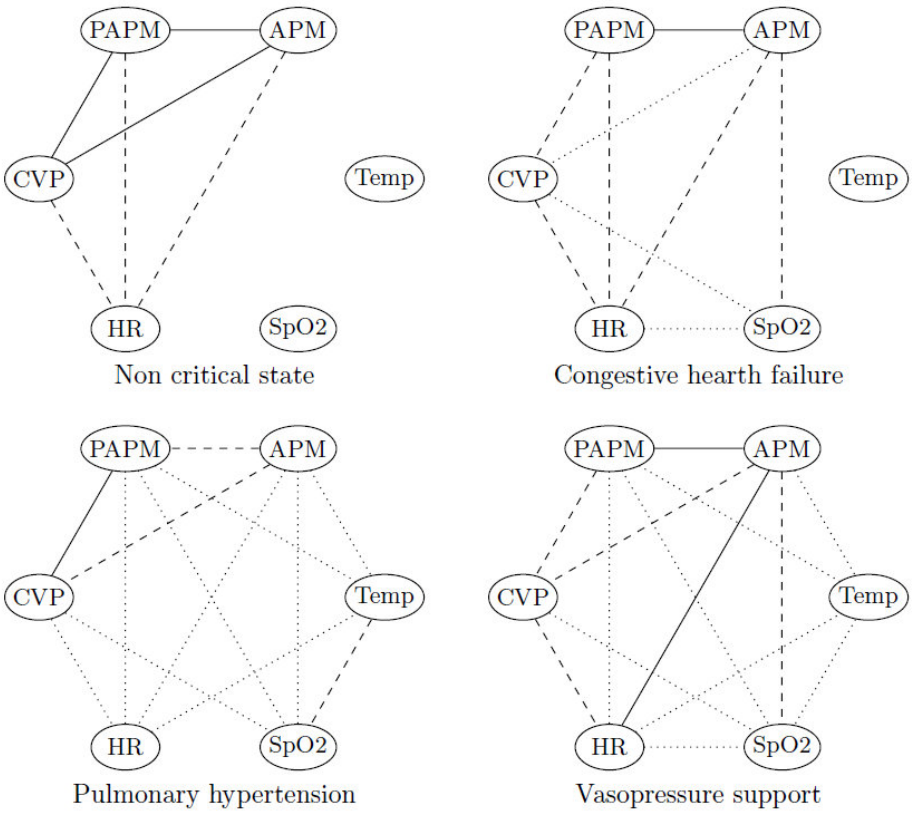


Figure 6.2:

For instance,*pulmonary hypertension* can be characterized by an elevated PAPM. Due to the increased right ventricular pressures we expect strong interactions between CVP and PAPM, but may attenuate the interactions between PAPM and APM as changes in PAPM will have a less than normal effect on left ventricular preload. The expected associations of vital signs show a different graph for the state of pulmonary hypertension than the relationships under normal physiological conditions.
For the clinical status of *congestive heart failure*, we find high correlation between APM and PAPM. This can be explained by a failure of the left ventricle, that leads to a decrease in APM, and the concurrent backward failure to an increase in PAPM.
For the status of *vasopressor support* there are strong correlations between APM and PAPM, too. But there are also higher partial correlations between HR and APM. This is due to the therapy which inhibits the normal autoregulation of the circulatory system. Hence, there are strong positive interactions between APM and PAPM, while the influence of CVP on the other variables is reduced.

In conclusion, we find that graphical models can detect physiological relationships between hemodynamic variables and the insights gained by this method are useful to improve online monitoring of vital signs.
In view of the high sampling rates of modern equipment this method is useful to identify these complications online in time critical situations on the intensive care unit.
Graphical analysis support the analysis of correlations in multivariate physiological time series.

Bibliography

[1] R. Dahlhaus, *Graphical interaction models for multivariate time series*, *Metrika*, **51**, pp. 157-172, 2000.

[2] T. P. Speed, H. T. Kiiveri, *Gaussian Markov Distributions Over Finite Graphs*, *The Annals of Statistics* **Vol. 14**, pp. 138-150, 1986.

[3] E. Avventi, *Spectral Moment Problems: Generalizations, Implementation and Tuning*, *KTH Doctoral Thesis, Stockholm, Sweden* 2011.

[4] M. J. Roberts, *Signals and Systems*, *Mc Graw Hill Higher Education* 2004.

[5] Ronald L. Allen, Duncan W. Mills, *Signal Analysis. Time, Frequency, Scale and Structure*, *Wiley-Interscience* 2004.

[6] U. Gather, M. Imhoff and R. Fried, Graphical models for multivariate time series from intensive care monitoring, *Statistics in Medicine*, **21** (2002), 2685-2701.

[7] J. Songsiri, J. Dahl and L. Vandenberghe, Graphical models of autoregressive processes, *Cambridge University Press*, (2010).

[8] M. Imhoff, R. Fried and U. Gather, Detecting Relationships Between Physiological Variables Using Graphical Modeling, *AMIA Annu Symp Proc.*, (2002).