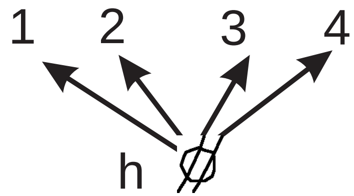
Special features of probability distributions: which are detectable by algebraic methods?

Nanny Wermuth, Chalmers/ University of Gothenburg, Sweden

Workshop 6-7 April 10, Warwick

A Gaussian one-factor analysis model

for 4 items (i.e. observed variables) is a linear system generated over



– where ⊘, the common parent node h, is a hidden variable

– for variables with zero mean and unit variance, let the simple correlation coefficients ρ_{ih}, for $i = 1, 2, 3, 4$, be positive

– the graph implies $\rho_{ik.h} = 0$ for each observed pair (i, k) so that $\rho_{ik} = \rho_{ih}\rho_{kh}$ and for observed $i \neq j \neq k \neq s$

$$\rho_{ik}/\rho_{jk} = \rho_{is}/\rho_{js},$$ the tetrad conditions

Early results on one-factor analysis models

– Bartlett (1951): if an observed covariance matrix $\boldsymbol{\Sigma}$ satisfies the tetrad conditions, so does its inverse

– Anderson and Rubin (1956): the tetrad conditions arise with a column vector $\boldsymbol{l}$, containing $\boldsymbol{\rho_{ih}}$, for $\boldsymbol{Y_i}$ observed, $\boldsymbol{Y_h}$ the hidden,

$$\boldsymbol{\Sigma = ll^T + \Delta, \quad \Delta \text{ diagonal}}$$

rank one of $\boldsymbol{ll^T}$ implies a zero determinant for each $\boldsymbol{2 \times 2}$ submatrix

recentlly: Drton, Sturmfels and Sullivan (2006), algebraic factor analysis

disappointing: no new insights regardinjg **improper representations**

Example of a positive definite correlation matrix, closed form for $\boldsymbol{l}$

$$\begin{pmatrix} \mathbf{1} & \mathbf{.84} & \mathbf{.60} \\ \mathbf{.} & \mathbf{1} & \mathbf{.38} \\ \mathbf{.} & \mathbf{.} & \mathbf{1} \end{pmatrix} \quad \boldsymbol{l} = \begin{pmatrix} \mathbf{1.2} \\ \mathbf{.7} \\ \mathbf{.5} \end{pmatrix} \boldsymbol{\Delta} = \begin{pmatrix} \mathbf{-.44} & \mathbf{.0} & \mathbf{0} \\ \mathbf{.} & \mathbf{.51} & \mathbf{0} \\ \mathbf{.} & \mathbf{.} & \mathbf{.75} \end{pmatrix}$$

improper because of the negative residual ‘variance’ in $\boldsymbol{\Delta}$

A special family of distributions for p symmetric binary variables $A_s, s = 1, ..., p$ i.e. **each has two equally probable levels**, coded as 1 and -1

We write e.g.

$$\pi_{111}^{A_1 A_2 A_3} = \Pr(A_1 = 1, A_2 = 1, A_3 = 1)$$

$$\pi_{1|11}^{A_1|A_2 A_3} = \pi_{111}^{A_1 A_2 A_3} / \pi_{11}^{A_2 A_3}$$

Covariance matrix is identical to the correlation matrix $\mathbf{P}$ with $\mathbf{H}$ upper-triangular, $\mathbf{\Delta}$ diagonal: triangular decomposition

$$\mathbf{P}^{-1} = \mathbf{H}^{\mathrm{T}} \mathbf{\Delta}^{-1} \mathbf{H}$$

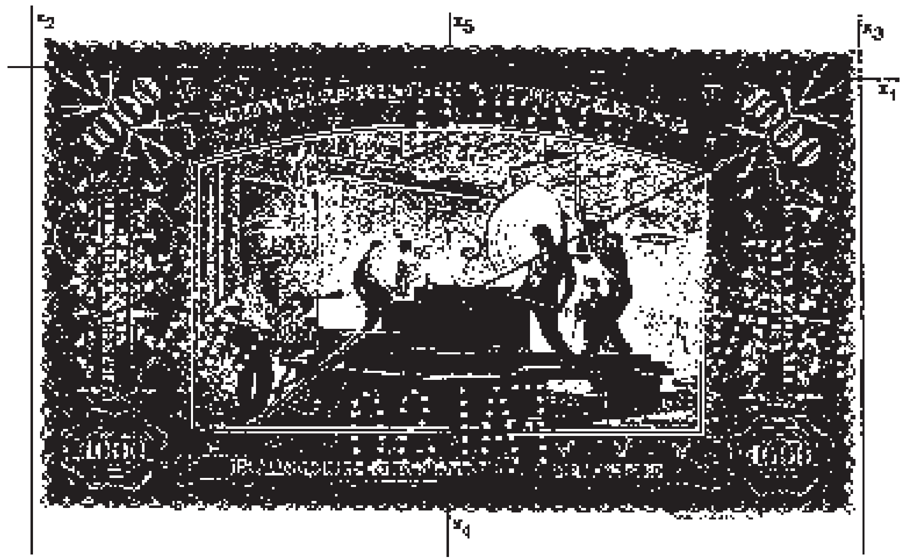
The linear triangular system of exclusively main effects in four variables is (Wermuth, Marchetti and Cox, 2009)

$$
\begin{aligned}
\pi_{i|jkl}^{A_1|A_2A_3A_4} &= \tfrac{1}{2}(1 + \eta_{12}ij + \eta_{13}ik + \eta_{14}il) \\
\pi_{j|kl}^{A_2|A_3A_4} &= \tfrac{1}{2}(1 + \eta_{23}jk + \eta_{24}jl) \\
\pi_{k|l}^{A_3|A_4} &= \tfrac{1}{2}(1 + \eta_{34}kl) \\
\pi_{l}^{A_4} &= \tfrac{1}{2}
\end{aligned}
$$

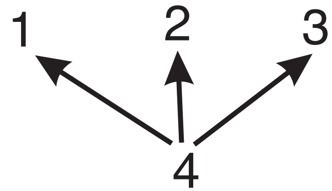
η's are linear regression coefficients of the binary variables

Example: 200 swiss bank notes of Riedwyl and Flury (1983)

Median-dichotomized values of $\boldsymbol{A_1}$: 145-length of the diagonal, $\boldsymbol{A_2}$: average distance of inner frame to the lower and upper border, $\boldsymbol{A_3}$: average height of the bank note, measured on the left and right; $\boldsymbol{A_4}$: real and forged



Mutual conditional independence of A_1, A_2, A_3 given A_4



The matrix **H** and the correlation matrix **P** $= \mathbf{H}^{-1}\boldsymbol{\Delta}\mathbf{H}^{-T}$ are

$$\mathbf{P} = \begin{pmatrix} 1 & \rho_{14}\rho_{24} & \rho_{14}\,\rho_{34} & \rho_{14} \\ . & 1 & \rho_{24}\rho_{34} & \rho_{24} \\ . & . & 1 & \rho_{34} \\ . & . & . & 1 \end{pmatrix} \quad \mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & -\rho_{14} \\ & 1 & 0 & -\rho_{24} \\ & & 1 & -\rho_{34} \\ 0 & & & 1 \end{pmatrix}$$

$\delta_{ss} = 1 - \rho_{s4}^2$ for $s = 1, 2, 3$

For the Swiss banknote data –

observed correlation matrix $\hat{\mathrm{P}}$ shows an almost perfect fit to $\mathbf{1} \perp\!\!\!\perp \mathbf{2} \perp\!\!\!\perp \mathbf{3} | \mathbf{4}$

– marginalising over $\boldsymbol{A_4}$ introduces strong associations

$$\hat{\mathrm{P}} = \begin{pmatrix} \mathbf{1} & \mathbf{0.92} & \mathbf{0.51} & \mathbf{0.98} \\ \mathbf{.} & \mathbf{1} & \mathbf{0.49} & \mathbf{0.95} \\ \mathbf{.} & \mathbf{.} & \mathbf{1} & \mathbf{0.51} \\ \mathbf{.} & \mathbf{.} & \mathbf{.} & \mathbf{1} \end{pmatrix} \quad \hat{\mathrm{H}} = \begin{pmatrix} \mathbf{1} & \mathbf{0.01} & \mathbf{-0.01} & \mathbf{-0.98} \\ & \mathbf{1} & \mathbf{-0.01} & \mathbf{-0.93} \\ & & \mathbf{1} & \mathbf{-0.51} \\ \mathbf{0} & & & \mathbf{1} \end{pmatrix}$$

and $\hat{\boldsymbol{\pi}}_{\mathbf{1111}} = \hat{\boldsymbol{\pi}}_{\mathbf{-1-1-1-1}} = \mathbf{.4}$

[What can be learned for this distribution by some algebraic factor analysis ?]

The more general multivariate regression chains

Let $\{1, \ldots, p\} = (a, b, c, d)$

$$f = f_{a|bcd} f_{b|cd} f_{c|d} f_d$$

gives a factorisation corresponding to the joint or single responses within the chain components a, b, c, d

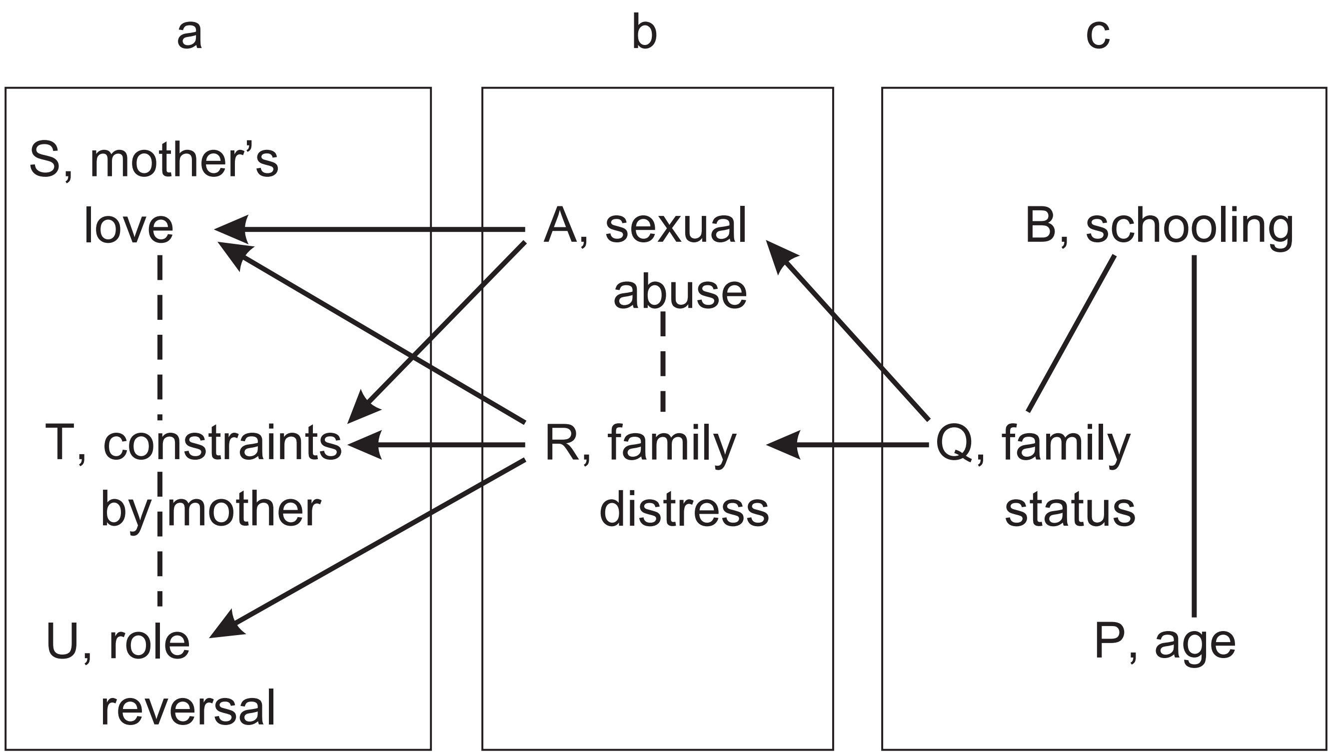
within each component: covariance graphs (**dashed lines**)

between components: regressions given the past (**arrows**)

[**within the last component**: a concentration graph (**full lines**)]

see Cox and/or Wermuth (1993, 2004, 2010), Drton (2009) Marchetti and Luparelli (2010), Kang and Tian (2009)

Childhood recollections of 283 healthy adult females



with: $\boldsymbol{a} \perp\!\!\!\perp \boldsymbol{c} | \boldsymbol{b}$ and $\boldsymbol{S} \perp\!\!\!\perp \boldsymbol{U} | \boldsymbol{A}, \boldsymbol{R}$ and $\boldsymbol{Q} \perp\!\!\!\perp \boldsymbol{P} | \boldsymbol{B}$

for the graph graph components: **directed acyclic** in blocks,

concentration graph within last block, **covariance graphs** within others

Special case: a fully recursive generating process with the independence structure captured by a **directed ayclic graph** means:

for the ordered node set $V = (1, 2, \ldots, d)$ and variable Y_i corresponding to node i, to generate the joint density

start with f_d

generate $f_{d-1|d}$

generate $f_{d-2|d-1,d}$

$\vdots$

generate $f_{1|2,\ldots,d}$

with univariate conditional densities of almost arbitrary form

We want to predict changes in structure when some variables are ignored and/or subpopulations are studied

which independences are preserved? when are dependences introduced? for which generating dependences are distortions introduced?

Most important needed property of the generated f_V :

an edge-inducing path is also association-inducing

path: sequence of edges coupling distinct nodes; **collider c:** $\circ \!\longrightarrow\!\!\succ c \prec\!\!\longleftarrow\! \circ$

inner nodes of a path: nodes of a path except for the endpoints

descendant i of k: a path of arrows starting from k, leading to i

$M = \{ \not\!\!\!\bigcirc \}$: marginalising set; $C = \{ \boxed{\bigcirc} \}$: conditioning set

Adapted from **Pearl (1988):** Let $\{a, b, M, C\}$ partition node set V.
A path from a to b in G^V_{par} is edge-inducing, iff every inner collider is in C or has a descendant in C and every other inner node is in M

Distributions with edge-inducing paths that are not association inducing

In the following $2 \times 2 \times 3$ table (Birch, 1963): $U \perp\!\!\!\perp V \mid W$ and $U \perp\!\!\!\perp V$

$28\pi_{uvw}$

	$w=1$		$w=2$		$w=3$	
	$v=1$	$v=2$	$v=1$	$v=2$	$v=1$	$v=2$
$u=1$	4	2	2	1	1	4
$u=2$	2	1	4	2	1	4
c. odds-r.	1		1		1	

with $\sum_w \pi_{u+w}\, \pi_{+vw}/\pi_{++w} = \pi_{u++} p_{+v+} = 1/4$

A family for $2 \times 2 \times 4$ tables with $U \perp\!\!\!\perp V \mid W$ and $U \perp\!\!\!\perp V$ with edge-inducing paths **that are not** association inducing (Studený 2002)

i.e.

$$U \prec\!\!-\!\!- \not\!\not\!\varnothing \prec\!\!-\!\!- V$$

with U dependent on $W = \not\!\not\!\varnothing$ and W dependent on V

does not lead to

$U \prec\!\!-\!\!- V$ with U dependent on V

$4\pi_{uvw}, \quad 0 < \epsilon < 1/2, \quad 0 < \delta < 1/2$

	$w = 1$		$w = 2$	
u	$v = 1$	$v = 2$	$v = 1$	$v = 2$
1	$(1-\epsilon)(1-\delta)$	$\epsilon(1-\delta)$	$\delta(1-\epsilon)$	$(1-\epsilon)(1-\delta)$
2	$\delta(1-\epsilon)$	$\delta\epsilon$	$\delta\epsilon$	$\epsilon(1-\delta)$
cor	1		1	

	$w = 3$		$w = 4$	
u	$v = 1$	$v = 2$	$v = 1$	$v = 2$
1	$\epsilon(1-\delta)$	$\delta\epsilon$	$\delta\epsilon$	$\delta(1-\epsilon)$
2	$(1-\epsilon)(1-\delta)$	$\delta(1-\epsilon)$	$\epsilon(1-\delta)$	$(1-\epsilon)(1-\delta)$
cor	1		1	

Both, conditional and marginal independence for connected and edge-minimal graphs only in incomplete families of distributions:

A family of distributions is **complete** if a function is implied to be zero whenever it has zero expectation for all members of the family

in a complete family with density $f(y)$

$$\int g(y)f(y)\;dy = 0 \implies g(y) = 0 \text{ a.s.}$$

Lehmann and Scheffé (1955), Mandelbaum and Rüschendorf (1987)

[What has algebraic statistics to say about complete families?]

Other needed properties of the generated $\boldsymbol{f_V}$ for deriving consequences for dependences in marginal/conditional distributions

For $\boldsymbol{a, b, c, d}$ disjoint subsets of $\boldsymbol{V}$, the family of distributions of $\boldsymbol{Y_V}$ is to satisfy

(1) the **intersection property**:

$$\boldsymbol{a \perp\!\!\!\perp b | cd} \text{ and } \boldsymbol{a \perp\!\!\!\perp c | bd} \text{ imply } \boldsymbol{a \perp\!\!\!\perp bc | d}$$

(2) the **composition property**:

$$\boldsymbol{a \perp\!\!\!\perp b | d} \text{ and } \boldsymbol{a \perp\!\!\!\perp c | d} \text{ imply } \boldsymbol{a \perp\!\!\!\perp bc | d}$$

see Dawid (1979) , Pearl (1988), Studený (2005) for general discussions

Necessary and sufficient conditions for Gaussian and discrete distributions to satisfy the intersection property: San Martin, Mouchart and Rolin (2005)

they give an example of a family for a $2 \times 3 \times 3$ table without the intersection property and with the marginal 3×3 table

	$j = 1$	$j = 2$	$j = 3$
$i = 1$	q_1	q_2	0
$i = 2$	0	0	q_3
$i = 3$	0	0	q_4

containing information common to the two variables i.e. event $\{A_2 = 1\}$ is the same as event $\{A_3 \neq 3\}$

instead in the following 5×4 table of probabilities

	$j=1$	$j=2$	$j=3$	$j=4$
$i=1$	q_{11}	0	0	q_{14}
$i=2$	0	q_{22}	q_{23}	0
$i=3$	q_{31}	0	0	q_{34}
$i=4$	0	q_{42}	0	q_{44}
$i=5$	q_{51}	0	q_{53}	0

A_2 contains no information about A_3 since $q_{ij}q_{i'j} > 0$ for all j

an extension to conditional probabilities $q_{ij|k}$

implies that the usual assumption of positive distributions is too strong

In a multivariate regression chain without a concentration graph

– the **Markov structure is defined by a set of pairwise independence statements** associated with the missing edges; Kang and Tian (2009)

– for discrete variables **a special sequence of multivariate logistic regression parameters gives the composition and the intersection property**; Marchetti and Luparrelli (2010)

– for discrete variables, each model defines a **curved exponential family**; Drton (2009)

[what can be learned from algebraic statistics, say for binary variables about the intersection and the composition property?]

How should the generating process look like to assure the desired properties of f_V?

– use a directed acyclic graph with special properties, called a **parent graph**

– constrain the **types of univariate conditional distributions**

The parent graph G^V_{par} is

a directed acyclic graph in node set $V = (1, 2, \ldots, d)$ that is

– connected

– has one compatible full ordering of V attached

– is edge-minimal for f_V

for $i \prec\!\!\!-\, k$: k is a **parent of offspring** i; par_i: the **set of parents of** i

edge-minimality of G^V_{par}

$$f_{i|\mathrm{par}_i} \neq f_{i|\mathrm{par}_i \setminus l} \text{ for each } l \in \mathrm{par}_i$$

(defines a research hypothesis; see Wermuth and Lauritzen, 1989)

Constraints on the generating process for the families of density, f_V

we denote the past of i by $\mathrm{pst}_i = \{i+1, \ldots, d\}$

(1) proper random responses Y_i depend just on Y_{par}

$f_{i|\mathrm{pst}_i} = f_{i|\mathrm{par}_i}$ **for each** $i < d$ **is varying fully**

(2) no constraints on parameters in the future from the past, i.e.

parameters of $f_{i|\mathrm{par}_i}$ **variation independent of parameters in** f_{pst_i}

Consequences of these mild assumptions on the generating process

f_V

– satisfies the **intersection property**, **the composition property**

– is a family of densities of a **complete family** of distributions

– in G^{V}_{par} **every edge-inducing path is association-inducing**

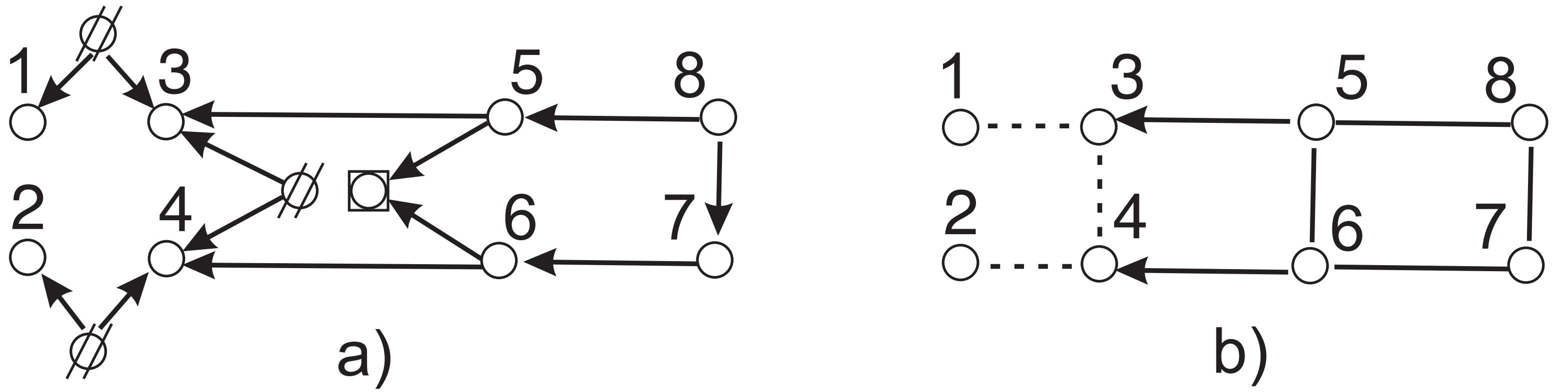
– a graph in node set $N = V \setminus C \cup M$ obtained by conditioning on $C = \{ \boxcircle \}$ and marginalizing over $M = \{ \oslash \}$,

G^{N}_{sum}, summarizes independences **and distortions in generating dependences** as implied by the generating process

A summary graph, $G^{N}_{\mathbf{sum}}$, with $N = V \setminus M \cup C$ is generated from a parent graph (or a multivariate regression graph or a summary in node set V by using a simple set of rules; see Wermuth (2010).

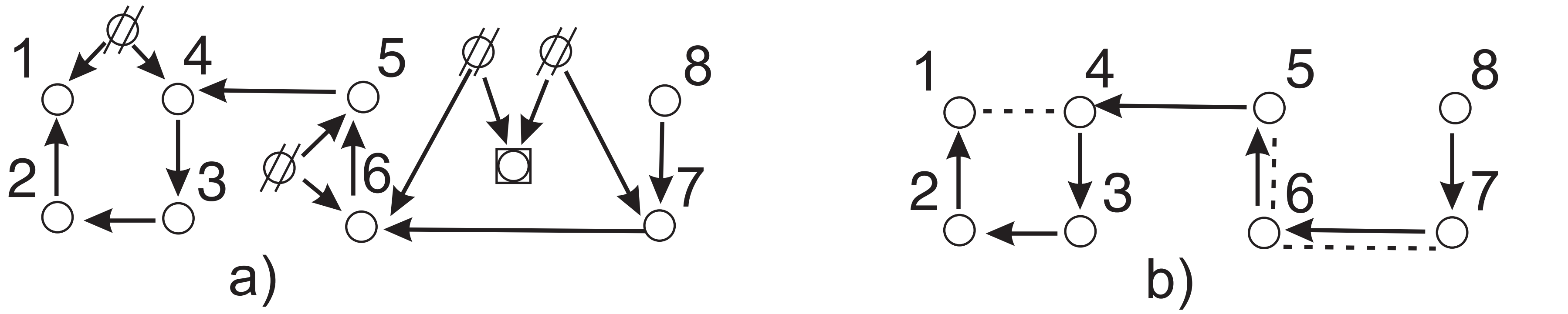
Example 1

A parent graph, a), that generates a multivariate regression chain graph, b)



Example 2

A parent graph, a), generating a summary graph with mixed directed cycles, b)



mixed directed cycles: the 4,4-path with inner nodes 1,2,3 and the 6,6-path via inner node 5 and the double edge for (6,7)

Multivariate regression chain graphs are summary graphs without mixed directed cycles

Summary

some of the outstanding features of multivariate regression chains that can have been generated over a larger parent graph

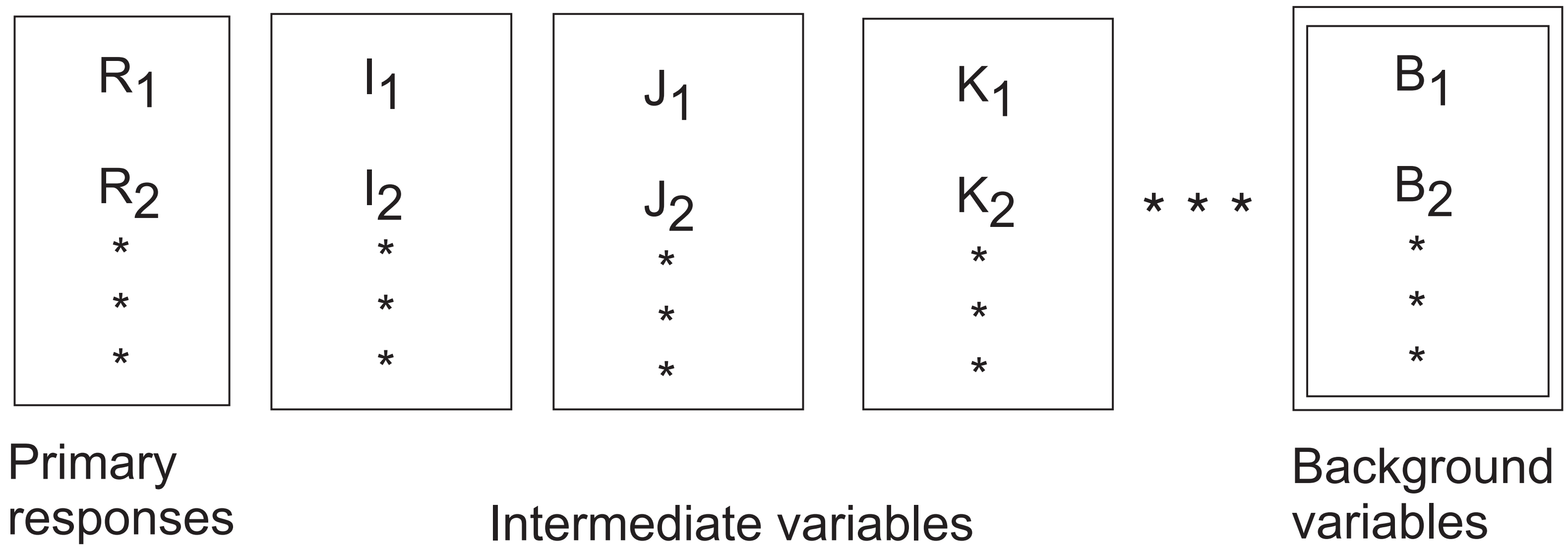
– pairwise independences define the Markov structure of the graph

– local modelling, flexibility regarding types of variable

– predicting changes in structure regarding independences **and** generating dependences with the summary graph.

Multivariate regression chains give a flexible tool for capturing development in observational studies and in controlled interventions

The general set-up



Conditioning **only on variables in the past**, i.e. variables on equal standing and in the future excluded; **with randomized interventions no direct dependences of hypothesized cause(s) on past variables**

Direct goals

we want to use the results to improve

– meta-analyses

– the planning of follow-up studies