Fields Institute, Toronto: September 2004

D.R.Cox

Nuffield College and Department of Statistics, Oxford

david.cox@nuf.ox.ac.uk

Joint work with

Nanny Wermuth

Chalmers/Gothenburg University, Sweden

wermuth@math.chalmers.se

Brief History

- Sewall Wright (1920); path analysis
- Tukey, Seneta (1950s)
- sociologists; Blalock, Duncan (1960-70)
- econometricians; Wold
- statisticians from early 1970s: Speed, Lauritzen, Wermuth
- computer scientists

Present emphasis: statistical analysis and interpretation

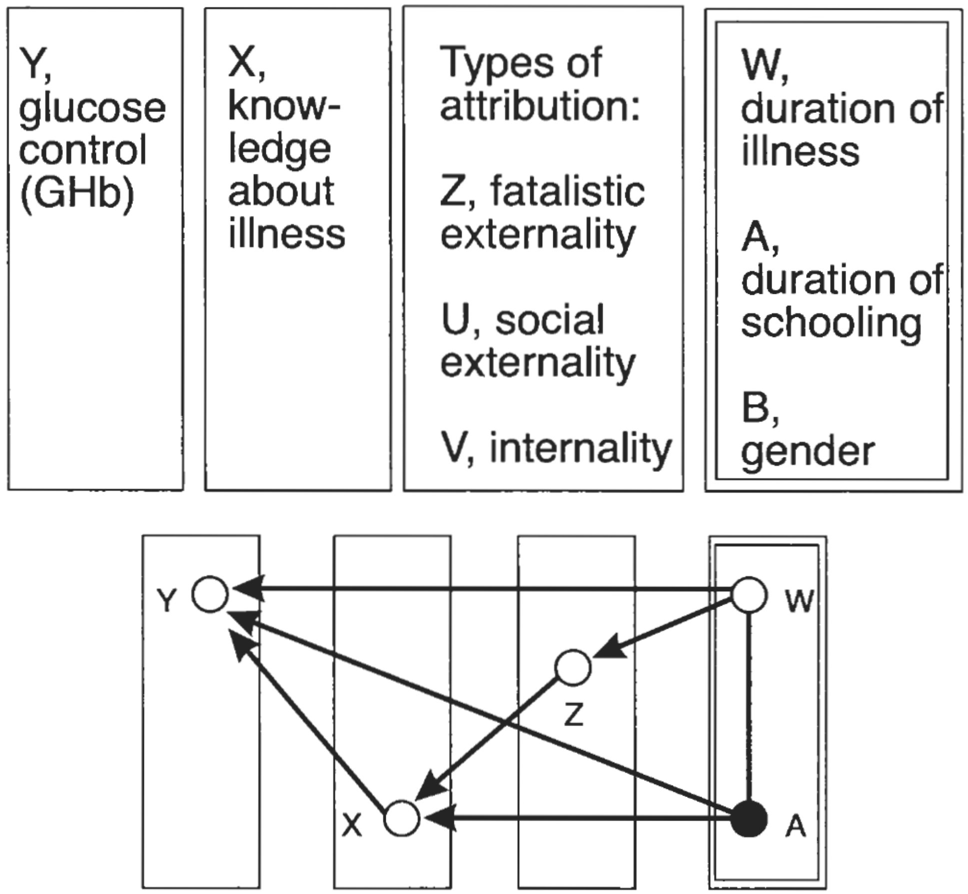
What determines the well-being of diabetic patients?

Y, glucose control (GHb)	X, know-ledge about illness	Types of attribution: Z, fatalistic externality U, social externality V, internality	W, duration of illness A, duration of schooling B, gender
primary response	intermediate variables		purely explanatory v.

6 quantitative variables: X, Y, Z, U, V, W, 2 binary variables: A, B;

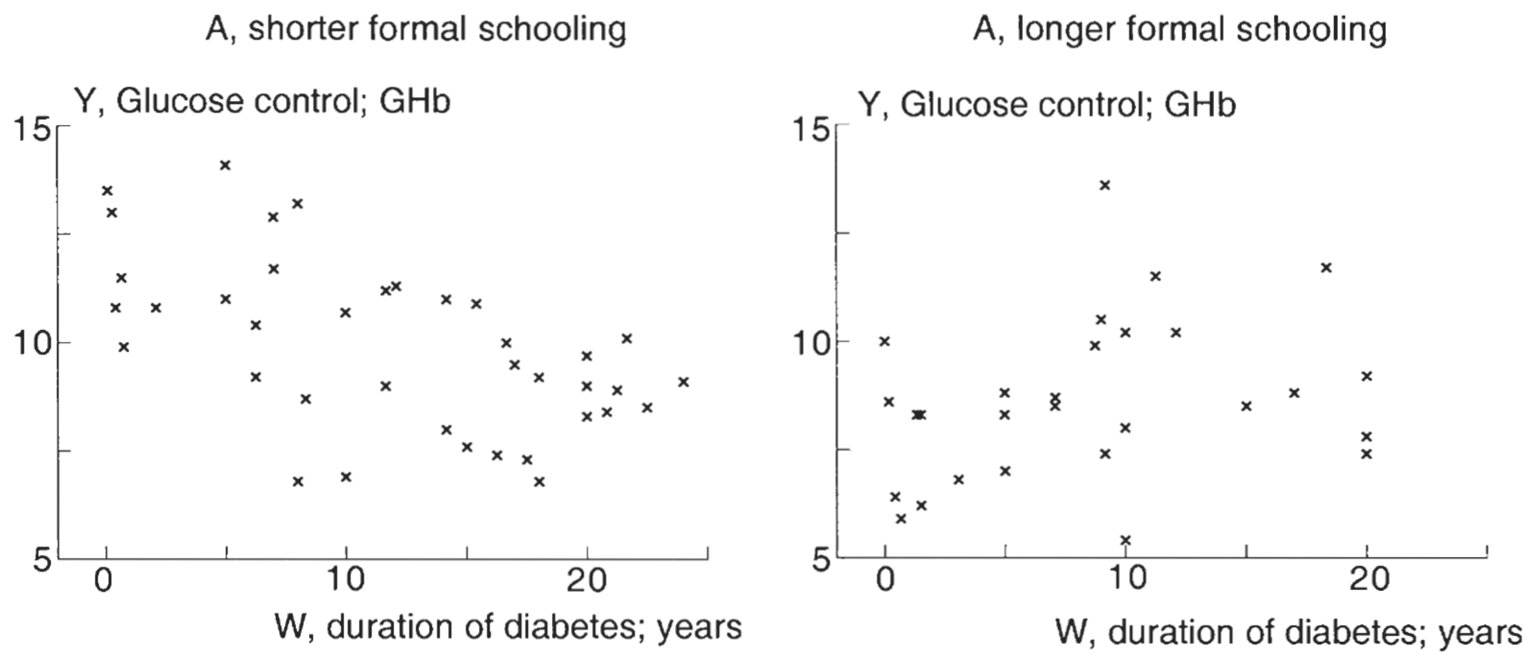
Cross-sectional study of 68 diabetic patients; 1990 Mainz

The graphical summary after analysis



$$Y: \; X + A * W, \quad X: \; Z + A, \quad Z: \; W.$$

The interactive effect $A * W$



Types of relation

For any two variables Y_i, Y_j

- either Y_i is a response to Y_j considered as explanatory (or *vice versa*
  - either by temporal ordering
  - or by subject-matter working hypothesis
- or Y_i and Y_j are on an equal footing
  - either because they are simply coordinates describing a single aspect
  - or because of a need to be agnostic about ordering

Some assumptions

Assume being on an equal footing is an equivalence relation, therefore sorting the variables into equivalence classes.

Within each class all variables are on an equal footing and pairs of variables in different classes have a response-explanatory relation.

Assume that relation consistent with the classes.

Graphical representation

Call the equivalence classes *blocks*.

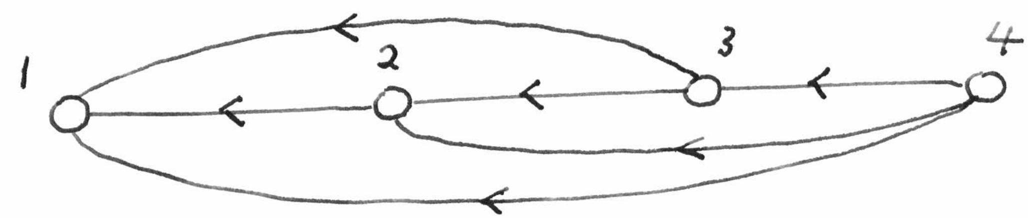
Assume blocks form an acyclic recursive system, starting with block d_B of purely explanatory variables to block 1 consisting of pure response or outcome variables.
The other variables are conveniently called *intermediate*.

Leave pairs of variables either joined by an edge or unjoined. Within block any edges are undirected. Between pairs of variables in different blocks any edge is directed (from explanatory to response). A missing edge represents some form of statistical independence.

Five canonical structures

Case 1: univariate recursive regression; fully directed graph

Complete graph



$$f_{1234} = f_{1|234} f_{2|34} f_{3|4} f_4$$

Linear form: $AY = \epsilon$, where $\mathrm{cov}(\epsilon) = \Delta$ is diagonal and A is upper triangular with unit diagonal elements:

$$a_{ij} = -\beta_{ij.i+1,\ldots,d\setminus j}$$

Further

$$\mathrm{cov}(Y) = \Sigma_{YY} = A^{-1}\Delta A^{-T}.$$

and

$$\mathrm{con}(Y) = \Sigma_{YY}^{-1} = A^{T}\Delta^{-1}A.$$

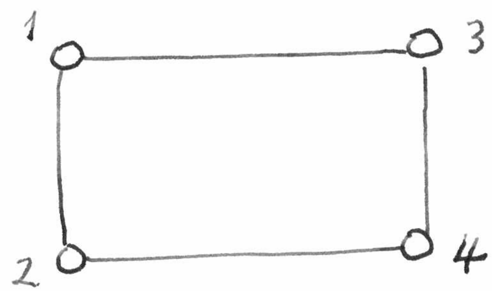
A special case; Markov chain

$$Y_1 \perp\!\!\!\perp Y_3, Y_4 \mid Y_2,\ f_{1234} = f_{1|2} f_{2|3} f_{3|4} f_4$$

A bidiagonal.

Characterize by $\mathcal{A}$ with zero elements where $a_{rs} = 0$ and unit elements elsewhere.

Case 2; Concentration graph: Chordless four cycle



$$1 \perp\!\!\!\perp 4 \mid 2, 3; 2 \perp\!\!\!\perp 3 \mid 1, 4$$

$$\Sigma_{YY}^{-1} = \mathrm{con}(Y)$$

has $\sigma^{14} = \sigma^{23} = 0.$

Characterize by an indicator matrix for Σ^{-1}.

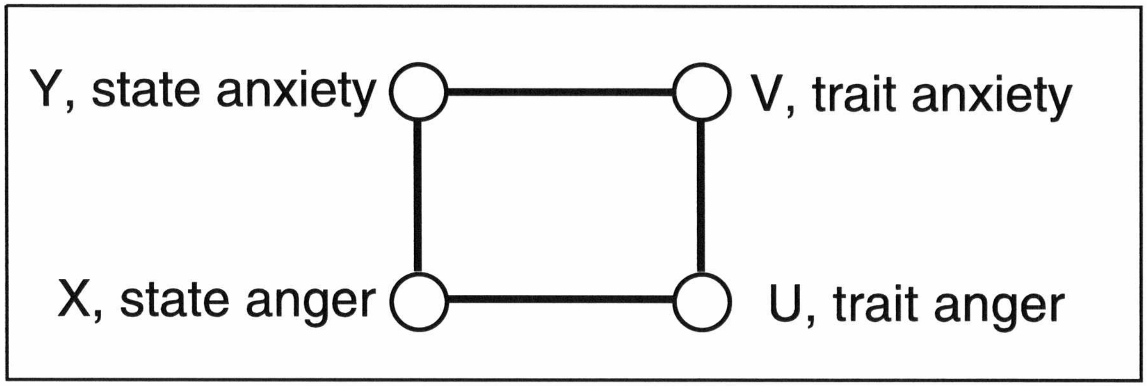
Examples of data with undirected graph structure

A Gaussian concentration graph model (Covariance selection)

Correlations among four psychological variables, 684 students. Marginal correlations (lower triangle), partial correlations given remaining (upper triangle).

	Y	X	V	U
Y, state anxiety	1	0.45	0.47	$-\mathbf{0.04}$
X, state anger	0.61	1	**0.03**	0.32
V, trait anxiety	0.62	0.47	1	0.32
U, trait anger	0.39	0.50	0.49	1

The corresponding graph (concentration graph) :



(The corresponding covariance graph is complete)

Case 3: Covariance graph

$$f_{14} = f_{23} = 0.$$

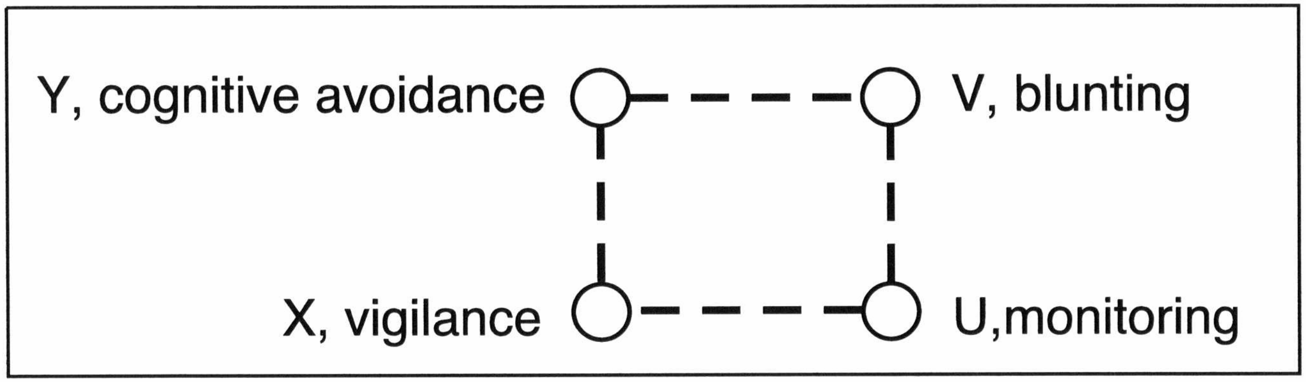
Covariance matrix Σ_{YY} has simplified structure specified by an indicator matrix.

A Gaussian covariance graph model (Linear in covariance structure)

Correlations among four strategies to cope with stress for 72 students. Marginal correlations (lower triangle), partial correlations given remaining (upper triangle).

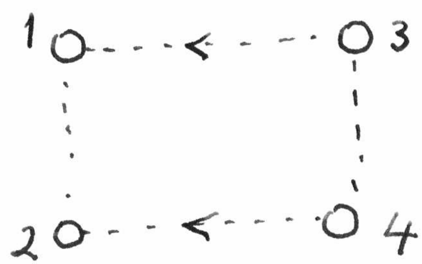
	Y	X	V	U
Y, cogn. avoid.	1	-0.30	0.49	0.21
X, vigilance	-0.20	1	0.21	0.51
V, blunting	0.46	**0.00**	1	-0.25
U, monitoring	**0.01**	0.47	-0.15	1

The corresponding graph covariance graph:



(The corresponding concentration graph is complete)

Case 4: Seemingly unrelated regression



$1 \perp\!\!\!\perp 4 \mid 3; 2 \perp\!\!\!\perp 3 \mid 4$

Can be checked from A but only after indirect calculation.

Data with a multivariate regression graph

Seemingly unrelated regressions

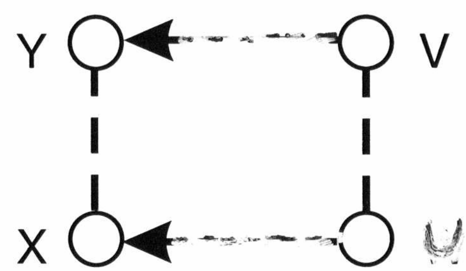
Correlations for blood-pressure variables, overweight, and age; 4 Systolic and diast. blood pressure averaged for several visits before a minor operation; 4 patients,

Variable	Y	X	V	U
Y, Log (syst/diast) bp	1	-.566	-.241	.300
X, Log diast. blood press.	-.544	1	-.107	.491
V, Weight(kg)/height(cm)	-.253	.336	1	.572
U, Age in years	-.131	.510	.608	1

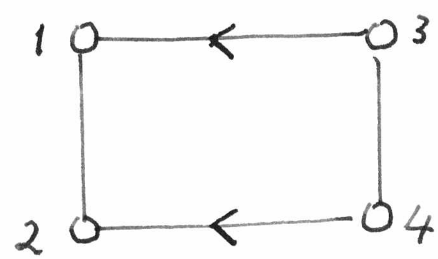
Standardized regression coefficients

$$\hat{\Pi}^*_{a|b} = \begin{pmatrix} \hat{\beta}^*_{yv.u} & \hat{\beta}^*_{yu.v} \\ \hat{\beta}^*_{xv.u} & \hat{\beta}^*_{xu.v} \end{pmatrix} = \begin{pmatrix} .486 & \mathbf{.040} \\ \mathbf{.037} & -.275 \end{pmatrix}$$

Corresponding joint response graph



Case 5: Block concentration regression



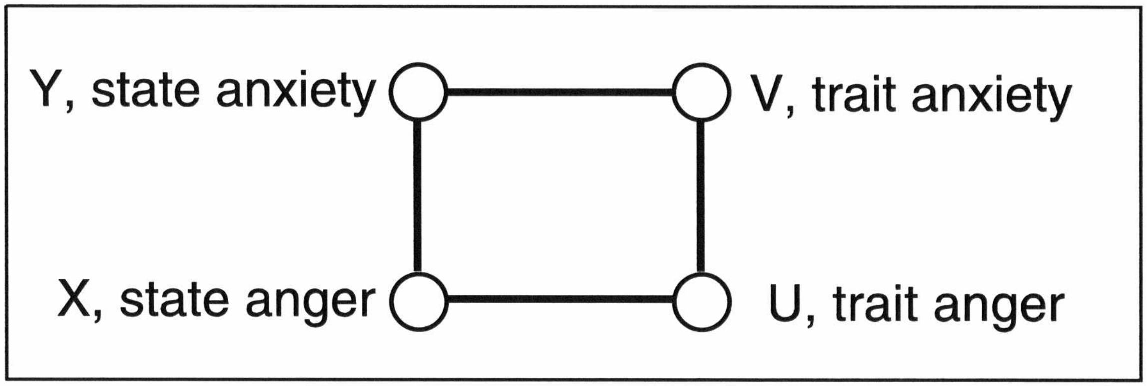
$$1 \perp\!\!\!\perp 4 \mid 2, 3 : 2 \perp\!\!\!\perp 3 \mid 1, 4$$

The matrix of regression coefficients

$$B_{(12)(34)} = \Sigma_{(12)(34)} \Sigma_{(34)}^{-1}$$

has zero elements.

The corresponding graph (concentration graph) :



(The corresponding covariance graph is complete)

Interpretation of chordless four-chain

Necessary and sufficient condition that an (undirected) concentration graph could have arisen from the dependencies in a univariate recursive regression is that there is no chordless k-chain for $k \geq 4$.

Condition for a model to define directly a possible data-generating process is that the specification can be used directly to simulate data with the right properties. Not applicable here even for Gaussian variables.

Can causality operate instantaneously?

Chordless four-chain

How can the chordless four-chain or its two block equivalent be generated?

Stochastic differential equation

$$\begin{pmatrix} Y_1(t+dt) \\ Y_2(t+dt) \end{pmatrix} = \begin{pmatrix} 1-\alpha_{11}dt & \alpha_{12}dt \\ \alpha_{21}dt & 1-\alpha_{22}dt \end{pmatrix} \begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix} + \begin{pmatrix} \beta_{11}dt & \beta_{12}dt \\ \beta_{21}dt & \beta_{22}dt \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \begin{pmatrix} dZ_1(t) \\ dZ_2(t) \end{pmatrix}.$$

Interpretation

Observe

- either single time of system in statistical equilibrium
- or time averages

Theoretical development

Many theoretical issues about the properties of and relations between the above systems, including the role of latent (unobserved) variables.

General independence properties can be tackled by graph-theoretical ideas leading for example to separation theorems or by matrix methods.

Gives explicit results for systems described by linear least squares regressions and in particular for Gaussian systems and, at the same time, often independence results for general systems based ultimately on univariate recursive regressions.

Univariate recursive regression

For each variable i we have the set of parents par_i. Then joint density is

$$f_Y = \Pi f_{i|\mathrm{par}_i}.$$

Linear system defined by

$$AY = \epsilon$$

with A upper triangular, unit diagonal elements and $a_{ij} = 0$ unless j is a parent of i.

For discussions of independence enough to consider $\mathcal{A} = \mathrm{In}(A)$ replacing all nonzero elements of A by one.

Some simple matrix properties

Because

$$Y = A^{-1}\epsilon$$

Covariance and concentration matrices of Y are

$$A^{-1}\Delta A^{-T}, A^T\Delta^{-1}A.$$

Thus undirected covariance and concentration graphs are determined by

$$\mathrm{In}(BB^T), \mathrm{In}(A^TA) = \mathrm{In}(\mathcal{A}^T\mathcal{A}),$$

where $B = A^{-1}$.

Indicator matrix of inverse

$$(2I - A)^{-1} = \{I - (A - I)\}^{-1} = I + (A - I) + (A - I)^2 + \ldots$$

Terms determine directed paths of various lengths from j to i.

Thus $\mathcal{B}$ is the ancestor matrix determining the ancestor graph of the system.

And so on to more complicated properties.

Wermuth, N. and Cox, D.R. (2004). *J.R.Statist. Soc.* B **66**, 687-717.

Summary

- relation between two variables
  - response- explanatory
  - on an equal footing
- directed and undirected edges
- two types of edge of each type
- rich variety of independency structures even for just four variables
- dual development via general distributions and via linear systems