E.
Causal effects in linear models

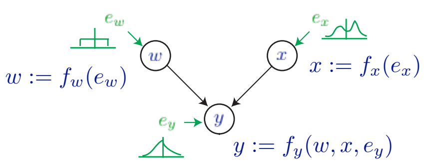
Structural Equation Models (SEM)

- Continuous-valued, linear special case of the non-parametric functional causal models considered in the previous section
- Commonly assumed (either explicitly or implicitly) that everything is Gaussian, so we can operate with means and covariance matrices only

  Note: This assumption does not generally mean that the method does not work when data is non-Gaussian, but rather that the methods are limited by what is possible for Gaussian data

- Practical when working with continuous-valued numerical data
- Used to a large extent in economics and social science
- Has received much criticism from statisticians. The original meaning of the model has been somewhat forgotten.

Reminder:
Functional causal model



Note:

- Values of e_w, e_x, e_y drawn from their respective distributions
- Values of w, x, y assigned based on the specified functions
- Specifies a joint distribution
- Directed acyclic graph!
- Some variables may be 'hidden' from us...

A linear Gaussian SEM:

A functional causal model, where...

- the variables are continuous-valued
- all functions $X_i := f_{x_i}(\mathrm{pa}(X_i))$ are linear, i.e. sums:

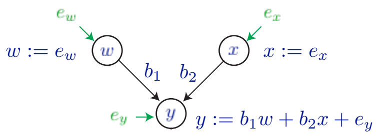
  $$X_i := \sum_{j \in \mathrm{pa}_i} b_{ij} X_j$$

  (may include constant terms, but left out here for simplicity)
- disturbances are Gaussian, which makes the joint distribution of the observed variables multidimensional Gaussian, all information in the covariance matrix

Note: SEMs based on acyclic graphs (DAGs) are called 'recursive' SEMs. To some degree people also use cyclic models ('non-recursive' SEMs), but the interpretation of such models is more difficult and also somewhat controversial, so we will not consider them but rather stick to models based on DAGs)

Example

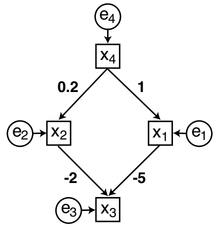
Linear gaussian structural equation model:

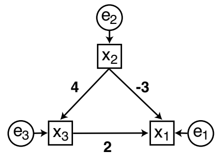


Note:

- Values of e_w, e_x, e_y drawn from Gaussian distributions
- Values of w, x, y assigned based on the linear functions
- Specifies a Gaussian joint distribution
- Directed acyclic graph!
- Some variables may be 'hidden' from us...

More examples:





$$x_4 := e_4$$
$$x_2 := 0.2x_4 + e_2$$
$$x_1 := x_4 + e_1$$
$$x_3 := -2x_2 - 5x_1 + e_3$$

$$x_2 := e_2$$
$$x_3 := 4x_2 + e_3$$
$$x_1 := -3x_2 + 2x_3 + e_1$$

- Alternative representation ('Causal Bayesian network') of a linear Gaussian SEM:
  - Same set of variables
  - Same DAG
  - Instead of 'disturbance variables' and deterministic functions, we have

$$P(x_i \mid \mathrm{pa}(x_i)) = \mathcal{N}\left(\mu_i, \sigma_i^2\right), \quad \mu_i = \sum_{j \in \mathrm{pa}(x_i)} b_{ij} x_j + c_i$$

Brief history of SEMs (à la Pearl)

- Developed in genetics (Wright 1921) and economics (Haavelmo 1943, Koopmans 1950,1953). The original users emphasized that
  - the graphs describe causal assumptions (prior knowledge)
  - the models are intended to aid decision making

  (Pearl, and this course, also strongly emphasizes these aspects!)
- Somewhere along the line the original ideas were forgotten, and slowly people began to think of SEMs mainly as a compact description of a covariance matrix...
  - a clear do()-notation was missing, people used = instead of :=
  - the connections between graphs and probability distributions were not yet understood

"I am speaking, of course, about the equation $\{y = a + bx + \epsilon\}$. What does it mean? The only meaning I have ever determined for such an equation is that it is a shorthand way of describing the conditional distribution of $\{y\}$ given $\{x\}$"

(Holland 1995, page 54)

SEM researchers: In an equation $\{y = a + bx + \epsilon\}$ in a SEM, when can we assign a causal meaning to b?

Pearl: Always! When an equation has been designated 'structural' then we have by definition that: $b = \frac{\partial}{\partial x} E\{y \mid \mathrm{do}(x)\}$

The appropriate question, rather, is: When is it possible to get consistent estimates of b and how does one construct such an estimator?

The covariance matrix implied by a SEM

- The data has been generated by
  $\mathbf{x} := \mathbf{B}\mathbf{x} + \mathbf{e}$
  where $\mathbf{B}$ contains the coefficients b_{ij} and if we arrange the variables in a causal order it is a lower-triangular matrix
- From this it follows, when no interventions are made, that...

$$\mathbf{x} = (\mathbf{I} - \mathbf{B})^{-1}\mathbf{e}$$

$$E\{\mathbf{x}\mathbf{x}^T\} = E\{(\mathbf{I} - \mathbf{B})^{-1}\mathbf{e}\mathbf{e}^T(\mathbf{I} - \mathbf{B})^{-T}\}$$

$$= (\mathbf{I} - \mathbf{B})^{-1}E\{\mathbf{e}\mathbf{e}^T\}(\mathbf{I} - \mathbf{B})^{-T} \quad \text{(in general)}$$

$$= (\mathbf{I} - \mathbf{B})^{-1}(\mathbf{I} - \mathbf{B})^{-T} \quad \text{(if disturbances have unit variance)}$$

- Note: $\mathbf{B}$ is the 'direct effects' matrix, whereas $\mathbf{A} = (\mathbf{I} - \mathbf{B})^{-1}$ gives the total effects. The latter can equivalently be calculated by multiplying and adding serial and parallel (respectively) effects

Identification of causal effects in linear Gaussian SEMs

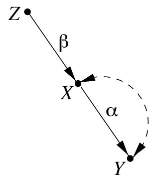
- Remember: a SEM is a special case of a general non-parametric DAG causal model, so if a causal effect is identifiable in the general case it must also be identifiable for a SEM!
- In addition, in some cases it will be possible to identify a causal effect in a SEM which is not identifiable in the general case.
- In a Gaussian SEM all do-probability distributions are Gaussian, so we only need to compute means and variances
- We can estimate effects using linear regression (see next slides)

- Calculating (total) causal effects in a fully specified SEM simply amounts to calculating $\mathbf{A} = (\mathbf{I} - \mathbf{B})^{-1}$.
- Direct effects can be read off from the graph (the coefficient directly gives the direct effect between the parent and the child)
- If we simply have the graph (but not the coefficients) and some uncontrolled data, then we can estimate the (total) causal effect of variable x on variable y as $r_{yx\cdot\mathbf{z}}$ where $\mathbf{z}$ is a set of variables that satisfies the back-door criterion with respect to (X, Y). [Pearl theorem 5.3.2]
- More generally, in order to identify any partial effect, as defined by a select bundle of causal paths from X to Y, we ought to find a set $\mathbf{Z}$ of measured variables that block all nonselected paths between X and Y. The partial effect will then equal the regression coefficient $r_{yx\cdot\mathbf{z}}$

In some cases, we can get stronger identification results for SEMs than for general non-parametric models...

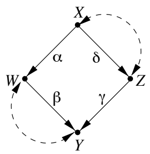
- 'Instrumental variable' formula

  Observing only X and Y we cannot estimate α. But additionally observing Z we can estimate it as $\alpha = r_{yz}/r_{xz}$



- Example:
  In the model on the right, we can identify α, β, and γ:

  $\alpha = r_{wx}$
  $\alpha\beta = r_{yx \cdot z}$
  $\Rightarrow \beta = r_{yx \cdot z}/r_{wx}$
  $\gamma = r_{yz \cdot x}$



Required reading

- This week's required readings are:
  - Parts 1 and 3 of chapter 5 of Pearl's book:
    http://bayes.cs.ucla.edu/BOOK-2K/ch5-1.pdf
    http://bayes.cs.ucla.edu/BOOK-2K/ch5-3.pdf
  - Pages 1-9 of Spirtes: 'The Limits of Causal Inference from Observational Data'
    http://www.hss.cmu.edu/philosophy/spirtes/rosenbaum.ps
    (note: link to pdf version available on the course homepage)