CompSci 295, Causal Inference

Rina Dechter, UCI

Lecture 5: Linear Causal Models

Causal Inference in Statistics, A primer, J. Pearl, M Glymur and N. Jewell Ch1, Why, ch1

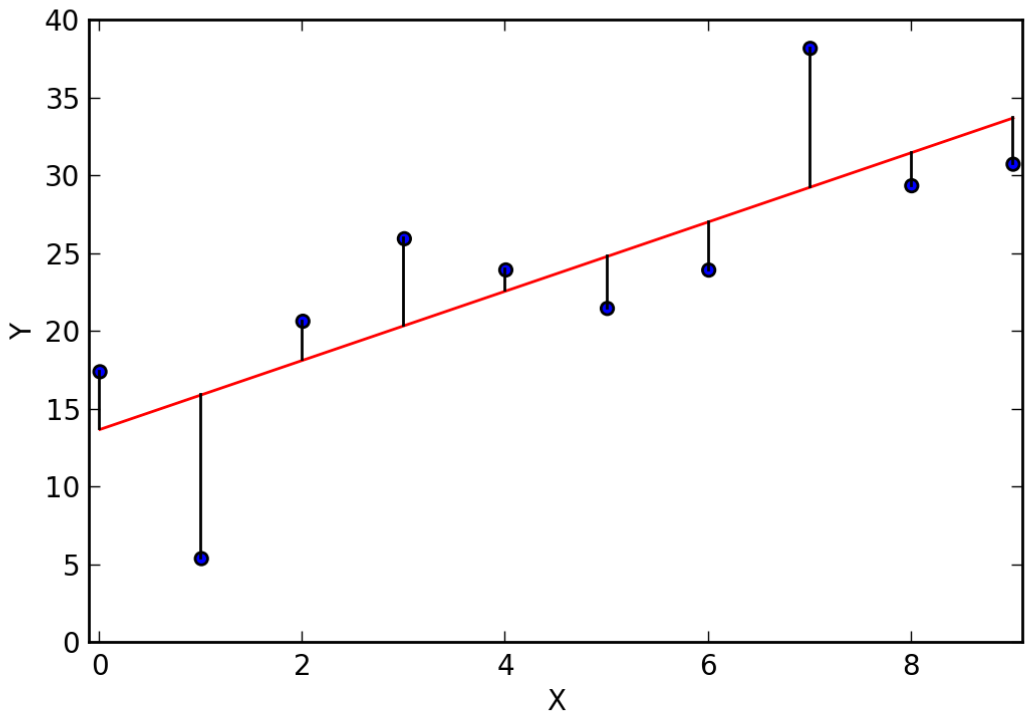
Outline

- Linear Regression
- Causal Inference in Linear Systems
- Structure vs Regression Coefficients
- The causal Interpretation of structural coefficients
- Identifying structural coefficients and causal effects
- Mediations

Outline

- Linear Regression
- Causal Inference in Linear Systems
- Structure vs Regression Coefficients
- The causal Interpretation of structural coefficients
- Identifying structural coefficients and causal effects
- Mediations

Regression



- Predict the value of Y based on X
- Used in Machine Learning too
- How to create a regression line?
    - Plot data values of X, Y
    - "Fit" them to y = mx + b
    - The least square regression is the line that minimize the sum of the squared error average $\sum (y - b - mx)^2$
    - Need to find b and m
        - What do they represent on the graph?

Regression Coefficient

- R_{YX} is slope of regression line of Y on X
- $m = R_{YX} = \sigma_{XY}/\sigma_X^2$
  - $R_{YX} = R_{XY}$?
  - When is it?
- Slope gives correlation
  - Positive number → positive correlation
  - Negative number → negative correlation
  - Zero → independent or non-linear

$$\sigma_{XY} \triangleq E[(X - E(X))(Y - E(Y))]$$

The covariance σ_{XY} is often normalized to yield the *correlation coefficient*

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Multiple Regression

- $y = r_0 + r_1 \cdot x + r_2 \cdot z$
- How do we visualize?
  - 3d plane
- What happens if we hold x at a value?
  - $r_1 \cdot x$ becomes a constant
  - r_2 is now the 2d slope of slice along X-axis
- What happens if we hold z at a value?
  - $r_2 \cdot z$ becomes a constant
  - r_1 is now the 2d slope of slice along Z-axis

Partial Regression Coefficient

- Symbol for regression coefficient of Y on X?
  - R_{YX}
- Symbol for regression coefficient of Y on X when holding Z constant?
  - $R_{YX \cdot Z}$
  - Called **partial regression coefficient**
- What happens when R_{YX} is positive and $R_{YX \cdot Z}$ is negative?
- What are partial regression coefficients in $y = r_0 + r_1 \cdot x + r_2 \cdot z$?
  - r_1 and r_2

Outline

- Linear Regression
- Causal Inference in Linear Systems
- Structure vs Regression Coefficients
- The causal Interpretation of structural coefficients
- Identifying structural coefficients and causal effects
- Mediations

Outline (Primer 3.8)

- Moving from non-paremtric causal models, to linear systems which are very popular in the social and behavioral sciences.

In this section, we examine in depth what causal assumptions and implications look like in systems of linear equations and how graphical methods can help us answer causal questions posed in those systems. This will serve as both a reinforcement of the methods we applied in nonparametric models and as a useful aid for those hoping to apply causal inference specifically in the context of linear systems.

For instance, we might want to know the effect of birth control use on blood pressure after adjusting for confounders; the total effect of an after-school study program on test scores; the direct effect, unmediated by other variables, of the program on test scores; or the effect

- Assumptions: relationships are linear and error is Gausian

Causal Inference In Linear Systems

Examples:

- What is the effect of birth control use on blood pressure after adjusting for confounders; or the total effect of an after-school study program on test scores;
- What is the direct effect or the unmediated by other variables, of the program on test scores.
- What is the effect of enrollment in an optional work training program on future earnings, when enrollment and earnings are confounded by a common cause (e.g., motivation).

- **Continuous variables:** We need to model with continuous variables. These traditionally been formulated as linear equation models .

- We will assume linear functions and Normal distributions of errors .

Linear systems are useful because

1. Efficient representation
2. Substitutability of expectations for probabilities
3. Linearity of expectations
4. Invariance of regression coefficients

Multivariate Gaussian can be expressed with expectation ond covariance on pairs of variables at most.
Also conditional probability can be captures by conditional expectation

Assuming Normal Distributions

- $P(Y \mid X, Z) = P(X \mid Z)$ can be written as $E[Y \mid X, Z] = E[Y \mid Z]$ (where Z is a set of variables).
- Linearity of expectation:
- Linearity permits fully specifying functions by path parameters (structure coefficient) along each edge
- Path coefficient are not regression coefficient.

$$E[Y|X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n] = r_0 + r_1 x_1 + r_2 x_2 + \cdots + r_n x_n$$

where each of the slopes $r1$, $r2$, . . . , rn is a partial regression coefficiet

Outline

- Linear Regression
- Causal Inference in Linear Systems
- Structure vs Regression Coefficients
- The causal Interpretation of structural coefficients
- Identifying structural coefficients and causal effects
- Mediations

Structrural vs Regression Coefficients

The path coefficient β along the edge $X \rightarrow Y$ quantifies the contribution of X in the function that defines Y in the model. For instance, if the function defines $Y = 3X + U$, the path coefficient of $X \rightarrow Y$ will be 3. The path coefficients $\beta_1, \beta_2, \ldots, \beta_n$ are fundamentally different from the regression coefficients $r_1, r_2, \ldots, r_n$ that we discussed in Section 1.3. The former are “structural” or “causal,” whereas the latter are statistical.

When we write $y = r_1 x + r_2 z + \epsilon$, as a regression equation, we are not saying that X and Z cause Y . We merely confess our need to know which values of r_1 and r_2 would make the equation $y = r_1 x + r_2 z$ the best linear approximation to the data, or, equivalently, the best linear approximation of $E(y|x, z)$.

We distinguish structural coefficients as α, β, and so on, and regression coefficients as r_1, r_2, and errors in regression equations are denoted ϵ_1, ϵ_2, and so on, as in Eq. (1.24),
and those in structural equations by U_1,U_2,

Structrural vs Regression Coefficients

The testable implications of nonparametric models are expressed as conditional independencies.

In linear models independencies are signified by vanishing correlation coefficients.
given the regression equation $y = r_0 + r_1 x_1 + r_2 x_2, \cdot\cdot\cdot, + r_n x_n + \epsilon$

if $r_i = 0$, then Y is independent of X_i conditional on all the other regression variables.

Outline

- Linear Regression
- Causal Inference in Linear Systems
- Structure vs Regression Coefficients
- The causal Interpretation of structural coefficients
- Identifying structural coefficients and causal effects
- Mediations

Controlled direct Effect in SEM

Definition: For any three variables X, Y , and Z where Z is a mediator between X and Y , the controlled direct effect (CDE) on Y of changing the value of X fromn$x$ to x' is defined as

$$CDE = P(Y = y|do(X = x), do(Z = z)) - P(Y = y|do(X = x'), do(Z = z)) \quad (3.18)$$

The Causal Interpretation of Structural Coefficients

It is easy to show that in a linear system, every path coefficient stands for the direct effect of the independent variable, X, on the dependent variable, Y

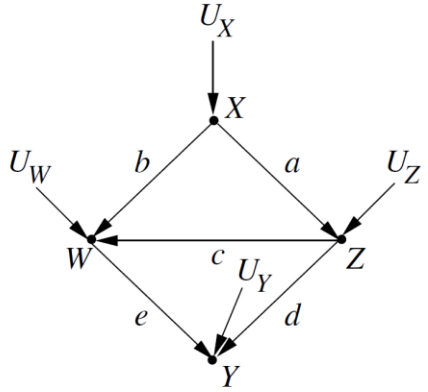
$CDE = P(Y = y|do(X = x), do(Z = z)) - P(Y = y|do(X = x'), do(Z = z))$

$$X = U_X$$

$$Z = aX + U_Z$$

$$W = bX + cZ + U_W$$

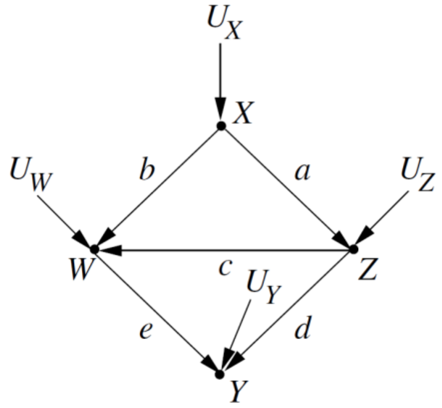
$$Y = dZ + eW + U_Y$$



In this example the controlled direct effect (CDE) of Z on Y becomes:
$DE = E[Y \mid do(Z = z + 1), do(W = w)] - E[Y \mid do(Z = z), do(W = w)]$
Since W is the only other parent of Y in the graph. What is the direct effect?

- Applying the do operators by deleting the appropriate equations from the model, the postincrease term in CDE becomes $d(z + 1) + ew$, and the preincrease term becomes $dz + ew$.
- The difference between the two is d, the path coefficient between Z and Y .

Direct vs Total Effect



- **Claim:** the total effect of X on Y is simply the sum of the products of the coefficients of the edges on every nonbackdoor path from X to Y . Or, to find the total effect of X on Y , first find every non-backdoor path from X to Y ; then, for each path, multiply all coefficients on the path together; then add up all the products.

- Why?

In our example: We want to find the total effect of Z on Y, we first intervene on Z, removing all arrows going into Z, then express Y in terms of Z in the remaining model.

$$Y = dZ + eW + U_Y$$
$$= dZ + e(bX + cZ) + U_Y + eU_W$$
$$= (d + ec)Z + ebX + U_Y + eU_W$$

The coefficient of Z is the total effect of Z on Y: the sum of the product of coefficients on the 2 nonbackdoor paths from Z to Y.

Outline

- Linear Regression
- Causal Inference in Linear Systems
- Structure vs Regression Coefficients
- The causal Interpretation of structural coefficients
- Identifying structural coefficients and causal effects
- Mediations

Identifying Structural Coefficients and Causal Effects.

- Identifiability: estimating the direct effects from non-experimental data

- It amounts to expressing the path coefficients associated with the total and direct effects in terms of the covariances σ_{XY} or regression coefficients $R_{YX \cdot Z}$, where X and Y are any 2 variables in the model and Z is a set of variables.

Identifying Structural Coefficients and Causal Effects.

The backdoor criterion gives us the set Z of variables we need to adjust for in order to determine the causal effect of X on Y .

We need only translate the backdoor procedure to the language of regression.

First: find a set of covariates Z that satisfies the backdoor criterion from X to Y in the model.
Then, regress Y on X and Z.

The coefficient of X in the resulting equation represents the true causal effect of X on Y .

Intuition: Because regressing on Z adds those variables into the equation, blocking all backdoor paths from X and Y , thus preventing the coefficient of X from absorbing the spurious information those paths contain.

Identifying Total Effect using the Backdoor Paths

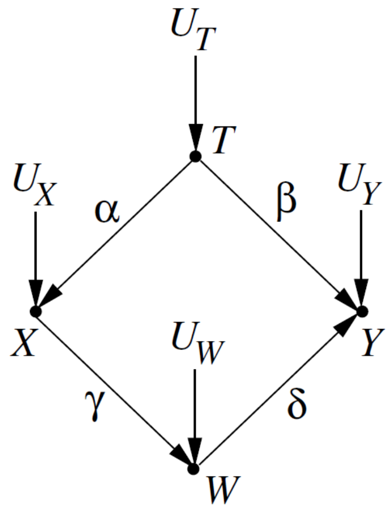
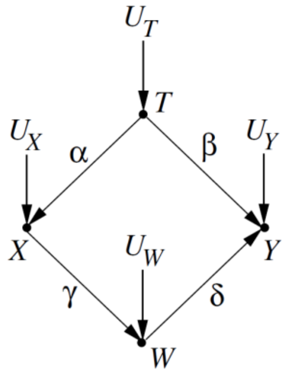


Figure 3.14: A graphical model in which X has no direct effect on Y, but a total effect that is determined by adjusting for T

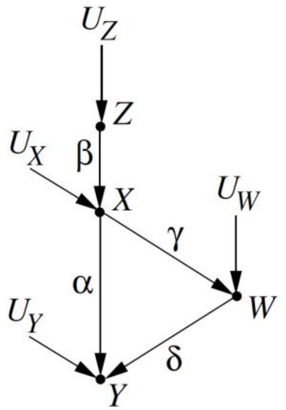
we must adjust for T, a backdoor . So we regress Y on X and T, using the regression equation $y = r_X X + r_T T + \epsilon$. The coefficient r_X is the total effect of X on Y .

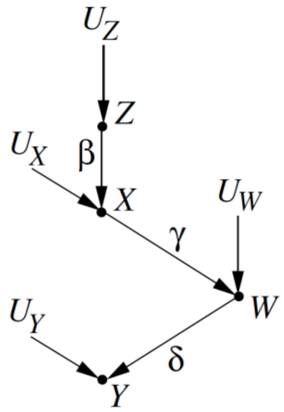
Notice that we don't need to look at any other variable. The graph permit us just to do that regression.

Finding Direct Effect



What is the direct effect of X on Y? …
0, in linear system it is the coefficient of X in the function for Y





Example: remove the edge between X and Y to get G_α
Here , W d-separates X and Y . So we regress Y on X and W, using the regression equation $Y = r_X X + r_W W + \epsilon$.
The coefficient r_X is the direct effect of X on Y .

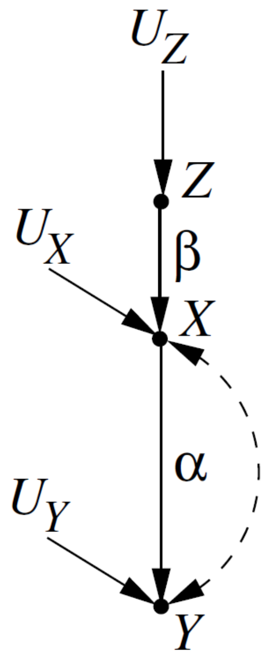
The Regression Rule for Identification :

- Use backdoor, but block also indirect paths going from X to Y
- First, remove edge from X to Y → G_α. If, in G_α, Z d-separates X and Y then regress Y on X and Z.
- The coefficient of r_X in the resulting equation is the structural coefficient α.

Summary on direct and total effects in SEM

- Regression is essential for identification and causal effect computation.
- To estimate causal effect we need to do a particular regression and specify:
  - What variables should be included
  - Which coefficient we are interested in.
- As long as we have a Markovian system every structural parameter can be identified this way. We can use various regression equations.
- But when some variables are not measurable or errors are correlated G_alpha can be used.

Instrumental Variable



- In this graph we cannot find the direct effect of X on Y via adjustment, because the dashed double-arrow arc. In this case, Z is an instrument with regard to the effect of X on Y that enables the identification of α

Definition: A variable is called an "instrument" if it is d-separated from Y in G_α and, it is d-connected to X.

Example: Z is an instrumental variable in the example.
We regress X and Y on Z separately, yielding

$$y = r_1 z + \epsilon \text{ , and } x = r_2 z + \epsilon.$$

Since Z emits no backdoors, r_2 equals β and r_1 equals the total effect of Z on Y , $\beta\alpha$.
Therefore, the ratio r_1/r_2 provides the desired coefficient α.

This example illustrates how direct effects can be identified from total effects but not the other way around.

Mediation in Linear Systems

In nonlinear systems, on the other hand, the direct effect is defined through expressions such as (3.18), or

$$DE = E[Y \mid do(x, z)] - E[Y \mid do(x', z)$$

where $Z = z$ represents a specific stratum of all other parents of Y (besides X).

Even when the identification conditions are satisfied, and we are able to reduce the do() operators (by adjustments) to ordinary conditional expectations, the result will still depend on the specific values of x, x', and z.

Moreover, the indirect effect cannot be given a definition in terms as do-expressions, since we cannot disable the capacity of Y to respond to X by holding variables constant. Nor can the indirect effect be defined as the difference between the total and direct effects, since differences do not faithfully reflect operations in non-linear systems.