Causal Inference with Graphical Models

Jose M. Peña
STIMA, IDA, LiU

Lecture 4: Linear-Gaussian Causal Models

Contents

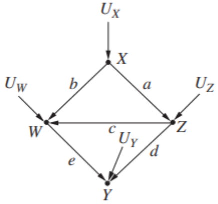
Literature

- Main sources
  - Pearl, J. *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge University Press, 2009. Chapter 5.
  - Pearl, J. Linear Models: A Useful "Microscope" for Causal Analysis. *Journal of Causal Inference*, 1:155-170, 2013.
- Additional sources
  - Pearl, J. Graphs, Causality, and Structural Equation Models. *Sociological Methods and Research*, 27:226-284, 1998.
  - Chen, B. and Pearl, J. Graphical Tools for Linear Structural Equation Modeling. Technical report R-432, UCLA, 2015.
  - Pearl, J. *Causality: Models, Reasoning, and Inference* (1st ed.). Cambridge University Press, 2000. Chapter 5.

Linear-Gaussian Causal Models

- Assume that
$$x_i = f_i(pa_i, u_i) = \sum_k \alpha_{ik} x_k + u_i$$
where $U \sim \mathcal{N}(0, \Sigma)$ st $\Sigma_{ij} = 0$ if $X_i \leftrightarrow X_j$ is not in the causal structure G, i.e. $U_i \perp_p U_j | \varnothing$.
- Note that $V \sim \mathcal{N}(0, (I - \alpha)^{-1} \Sigma (I - \alpha)^{-1})$.
- Assume without loss of generality that the variables are standardized to have zero mean and unit variance (it simplifies some expressions such as expectations and path analysis).
- The normality assumption may be explained via the central limit theorem by letting U_i represent the sum of many iid unobserved microprocesses.
- The linearity assumption promotes interpretability by allowing to annotate G with the **path coefficients** α_{ik}.

$$X = U_X$$
$$Z = aX + U_Z$$
$$W = bX + cZ + U_W$$
$$Y = dZ + eW + U_Y$$



Linear-Gaussian Causal Models

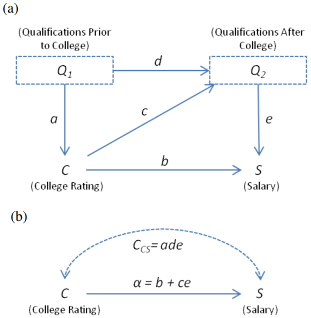


Figure 1. (a) Model with latent variables (Q_1 and Q_2) shown explicitly (b) Same model with latent variables summarized

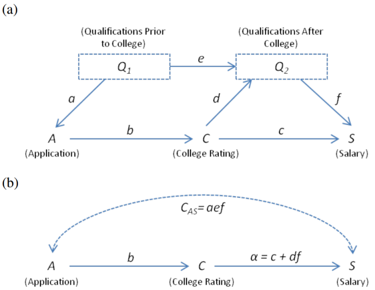
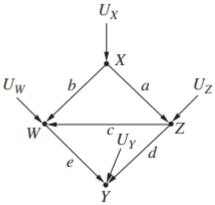


Figure 2. Graphs associated with Model 3 in the text (a) with latent variables shown explicitly (b) with latent variables summarized

Linear-Gaussian Causal Models

$$X = U_X$$
$$Z = aX + U_Z$$
$$W = bX + cZ + U_W$$
$$Y = dZ + eW + U_Y$$



- By modifying the appropriate equations,

$$CDE(Z,Y) = E[Y|do(z+1,w)] - E[Y|do(z,w)] = d(z+1) + ew - dz - ew = d.$$

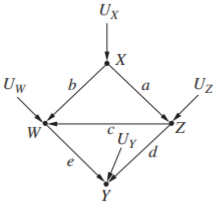
- Note that $NDE(Z,Y) = CDE(Z,Y) = DE(Z,Y)$ because the value at which W is fixed is irrelevant.
- Note also that

$$\frac{\partial}{\partial z} E[Y|do(z,w)] = d.$$

Linear-Gaussian Causal Models

- $TE(Z, Y)$ = **sum of the products** of the path coefficients of the edges on every directed path from Z to Y. In the example

$$X = U_X$$
$$Z = aX + U_Z$$
$$W = bX + cZ + U_W$$
$$Y = dZ + eW + U_Y$$



we have that

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

and, thus, an increase of one unit in Z will increase Y by $d + ec$ units.

- Note that $DE(z+1, z, Y) = -DE(z, z+1, Y)$ and, thus, $TE(z+1, z, Y) = IE(z+1, z, Y) - DE(z, z+1, Y) = IE(z+1, z, Y) + DE(z+1, z, Y)$.
- Then, $IE(Z, Y)$ = sum of the products of the path coefficients of the edges on every directed path from Z to Y with the **exception** of $Z \rightarrow Y$.
- Note that $TE(\{Z_1, \ldots, Z_m\}, \{Y_1, \ldots, Y_n\}) = (\sum_{i=1}^{m} TE(Z_i, Y_j))_{j=1,\ldots,n}$. Hence, we assume singletons hereinafter.

Multiple Linear Regression

- Linear regression: Predict $E[Y|x]$ where Y is the response or dependent variable and x is a value of the explanatory or independent variable X. Assume that $Y|x \sim \mathcal{N}(\mu(x), \sigma^2)$ and $E[Y|x] = \mu(x) = \beta_0 + \beta_X x$. It is known that
$$\beta_X = r_{YX} = \frac{\partial}{\partial x} E[Y|x] = \frac{\sigma_{XY}}{\sigma_X^2}.$$
- Multiple linear regression: Predict $E[Y|x_1, \ldots, x_n]$ under the assumption that $Y|x_1, \ldots, x_n \sim \mathcal{N}(\mu(x_1, \ldots, x_n), \sigma^2)$ and $E[Y|x_1, \ldots, x_n] = \mu(x_1, \ldots, x_n) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$. It is known that
$$\beta_i = r_{YX_i \cdot X_1, \ldots, X_{i-1}, X_{i+1}, \ldots X_n} = \frac{\partial}{\partial x_i} E[Y|x_1, \ldots, x_n] = \frac{\sigma_{YX_i \cdot X_1, \ldots, X_{i-1}, X_{i+1}, \ldots X_n}}{\sigma^2_{X_i \cdot X_1, \ldots, X_{i-1}, X_{i+1}, \ldots X_n}}.$$
- Note that $Y|x_1, \ldots, x_n \sim \mathcal{N}(\mu(x_1, \ldots, x_n), \sigma^2)$ is equivalent to $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n + \epsilon$ where $\epsilon \sim \mathcal{N}(0, \sigma^2)$.

Multiple Linear Regression

- Consider some samples $\{(x_{i1}, \ldots, x_{in}, y_i)\}_{i=1}^{N}$. The maximum likelihood estimates of the **regression coefficients** can be obtained by maximizing the function

$$p(\{y_i\}|\{(x_{i1}, \ldots, x_{in})\}) = \prod_i \mathcal{N}(y_i|\mu_i, \sigma) = \prod_i \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-\frac{1}{2\sigma^2}(y_i-\mu_i)^2}$$

  where $\mu_i = \mu(x_{i1}, \ldots, x_{in}) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_n x_{in}$.

- This is equivalent to minimizing the function

$$-\ln p(\{y_i\}|\{(x_{i1}, \ldots, x_{in})\}) = \sum_i \frac{1}{2\sigma^2}(y_i - \mu_i)^2 + \frac{N}{2}\ln\sigma^2 + \frac{N}{2}\ln 2\pi$$

  which implies that ML estimates coincide with the least squares estimates obtained by minimizing the function

$$\sum_i (y_i - \mu_i)^2.$$

- The LS minimization problem can be solved analytically as

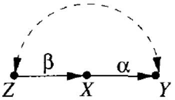
$$\hat{\boldsymbol{\beta}} = (\boldsymbol{x}^T\boldsymbol{x})^{-1}\boldsymbol{x}^T\boldsymbol{y}$$

  where $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \ldots, \hat{\beta}_n)^T$, $\boldsymbol{x} = (1, x_{i1}, \ldots, x_{in})_{i=1}^{N}$, and $\boldsymbol{y} = (y_1, \ldots, y_N)^T$.

Single-Door Criterion

- The path coefficient α associated with the edge $X \to Y$ (i.e., $DE(X, Y)$) is identifiable if there exists a set of variables Z st
  - Z contains no descendants of Y, and
  - Z blocks **every** path between X and Y with the **exception** of $X \to Y$.

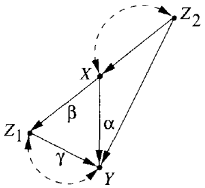
  Moreover, α coincides with the **regression coefficient** $r_{YX \cdot Z}$ (which is obtained by regressing Y on X and Z).



- Proof: Consider the **regression equation** $y = ax + bz + \epsilon$, and let $Z = 0$. Then, $y = ax + \epsilon$. Consider the **structural equation** $y = \alpha x + \gamma(pa_Y \setminus x) + u_Y$, and let $pa_Y \setminus x = 0$. Then, $y = \alpha x + u_Y$. Since $X \to Y$ is the only path between X and Y that is not blocked by Z, the regression and the structural equations coincide and, thus, $\alpha = a = r_{YX \cdot Z}$.
- In a causal structure with no latent variables, every path coefficient is identifiable by taking $Z = Pa_Y \setminus X$.

Back-Door Criterion

- The path coefficient α cannot be identified by the single-door criterion in

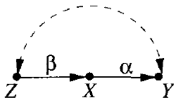


  but $TE(X, Y)$ is identifiable.
- $TE(X, Y)$ is identifiable if there exists a set of variables Z st
  - Z contains no descendants of X, and
  - Z blocks every path between X and Y that contains an arrow into X.

  Moreover, $TE(X, Y)$ coincides with the **regression coefficient** $r_{YX \cdot Z}$.
- Proof: Consider the **regression equation** $y = ax + bz + \epsilon$, and let $Z = 0$. Then, $y = ax + \epsilon$. Since Z blocks all the back-door paths from X to Y, the only unblocked paths are the directed ones and, thus, the **regression and the total effect equations** coincide and, thus, $TE(X, Y) = a = r_{YX \cdot Z}$.

Back-Door Criterion

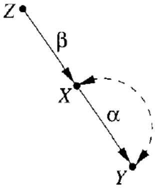
- The back-door criterion does not subsume the single-door criterion: $TE(Z,Y)$ cannot be identified by the back-door criterion in



  but $TE(Z,Y) = \alpha\beta$ and, thus, it can be identified by two applications of the single-door criterion, since $\alpha = r_{YX \cdot Z}$ and $\beta = r_{XZ}$.
- The single- and back-door criteria are special cases of a more general result: The **partial effect** of X on Y due to some directed paths from X to Y can be identified as $r_{YX \cdot Z}$ if Z blocks all the other paths between X and Y.

Instrumental Variables

- The path coefficient α cannot be identified by the methods above in

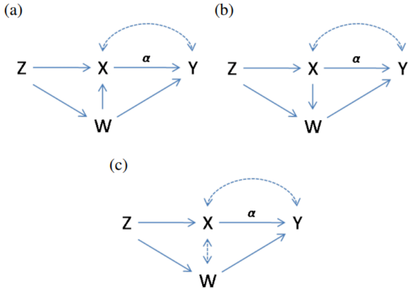


  but $\beta = r_{XZ}$ and $TE(Z, Y) = \alpha\beta = r_{YZ}$ and, thus, $\alpha = r_{YZ}/r_{XZ}$. The variable Z is called **instrumental variable**. Recall that this effect was unidentifiable in the nonparametric model.
- Intuitively, Z is an instrumental variable if Z causes Y and all correlation between Z and Y is mediated by X.

Instrumental Variables

- A variable Z is a conditional instrumental variable given a set W for the path coefficient α associated with the edge $X \to Y$ if
  - X is a descendant of Z,
  - W contains no descendants of Y, and
  - W blocks every path between Z and Y that does not contain $X \to Y$.

  Then, $\alpha = r_{YZ \cdot W} / r_{XZ \cdot W}$.



Is Z an instrument given W ? (a) Yes. (b) No. (c) Yes.

- Recall that α is not identifiable for non-parametric causal models. However, it is identifiable for linear-Gaussian causal models. So, *do*-calculus is sound but **not** complete for linear-Gaussian causal models.

Path Analysis

- The covariance (= correlation or regression coefficient in a standardized model) between two variables X and Y can be expressed as the **sum of the products** of the path coefficients and error covariances of the edges on every unblocked path between X and Y.
- This results in a system of equations that may allow to compute the path coefficients.

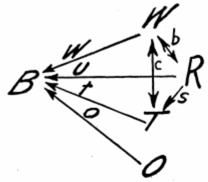


FIG. 14.—Relations between wet-bulb depression (B), wind velocity (W), radiation (R), and temperature (T) as assumed for direct analysis.

Six equations can be formed, expressing the six known correlations in terms of the unknown path coefficients. A seventh equation represents the complete determination of B by W, R, T, and O.

(1) $r_{BW} = \quad 0.28 = w + t(c+bs) + ub.$
(2) $r_{BR} = \quad .48 = wb + ts + u.$
(3) $r_{BT} = \quad .59 = w(c+bs) + t + us.$
(4) $r_{WR} = - \; .01 = b.$
(5) $r_{WT} = - \; .02 = c + bs.$
(6) $r_{RT} = \quad .47 = s.$
(7) $o^2 + w^2 + t^2 + u^2 + 2wt(c+bs) + 2wub + 2uts = 1.$

- Pearl's work owes much to path analysis (Wright, 1921), e.g. the model identifiability result in the last slide builds on Wright's equations.

Selection Bias

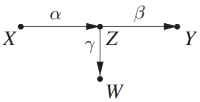


Figure 2 Path diagram depicting an intermediate variable (*Z*) and its proxy (*W*). Conditioning on *W* would distort the regression of *Y* on *X*.

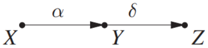


Figure 4 Conditioning on *Z*, a descendant of *Y*, biases the regression of *Y* on *X*.

- In standardized models, $r_{YX \cdot Z} = \frac{r_{YX} - r_{YZ} r_{ZX}}{1 - r_{ZX}^2}$.
- Path analysis in Figure 2: $r_{YX \cdot W} = \frac{\alpha\beta - \beta\gamma^2\alpha}{1 - \alpha^2\gamma^2} \neq \alpha\beta = TE(X, Y)$.
- Path analysis in Figure 4: $r_{YX \cdot Z} = \frac{\alpha - \alpha\delta^2}{1 - \alpha^2\delta^2} \neq \alpha = TE(X, Y)$.
- Selection bias: Samples are preferentially selected depending on the values of some variables in the model (e.g., $Z = z$ in Figure 4, and $W = w$ in Figure 2 which is a proxy of $Z = z$).
- It is like new paths get unblocked through the latent variables (e.g., $X \to Z \leftarrow U_Z \to Z \to Y$ in Figure 2, and $X \to Y \leftarrow U_Y \to Y$ in Figure 4, a.k.a. virtual colliders).
- Note that the back-door criterion warns against these adjustments.

Model Identifiability

- If a causal model has no bow (i.e., no subgraph of the form $A \to B \leftrightarrow A$), then the model is identifiable with probability almost 1. In other words, the matrices α and Σ can be determined uniquely from the population covariance matrix $(I - \alpha)^{-1}\Sigma(I - \alpha)^{-1}$.
- This criterion subsumes the single-door criterion for model identification.
- Consider the bow causal model

$$x = u_X$$

$$y = \alpha x + u_Y$$

  with $U_X \not\perp U_Y | \varnothing$. Then, path analysis implies that $\sigma_{XY} = \alpha + \sigma_{U_X U_Y}$, which implies that the model (i.e., α) is unidentifiable.
- Model unidentifiability does not necessarily imply unidentifiability of every causal effect, as shown in previous examples. Therefore, the single-door criterion may still be useful when the model is not identifiable.

Summary

- Linear-Gaussian Causal Models
- Multiple Linear Regression
- Single-Door Criterion
- Back-Door Criterion
- Instrumental Variables
- Path Analysis
- Selection Bias
- Model Identifiability

Thank you