Machine Learning in Bioinformatics

Causal Thinking

Philipp Benner
philipp.benner@bam.de

VP.1 - eScience
Federal Institute of Materials Research and Testing (BAM)

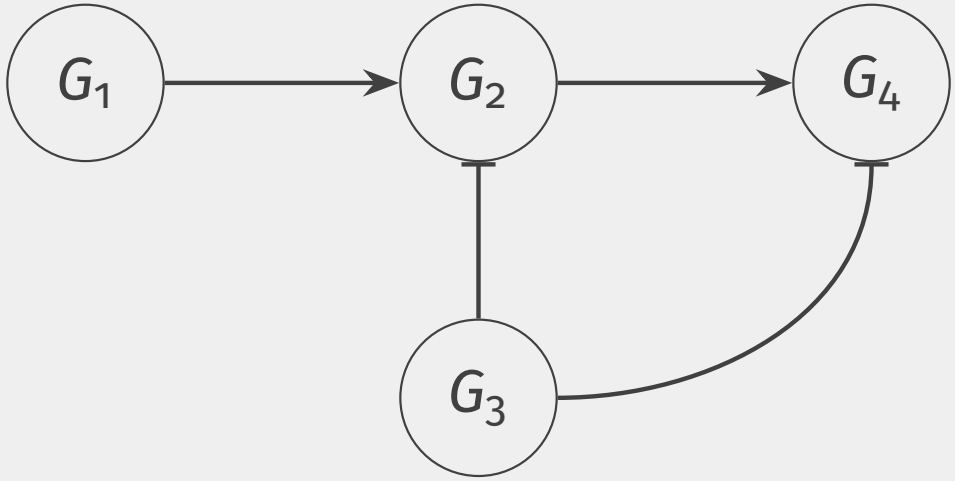
February 8, 2026

Motivation

Gut microbiota cause disease

- Gut microbiota, gut microbiome, or gut flora are the microorganisms, including bacteria, archaea, fungi, and viruses, that live in the digestive tracts
- Variations in the gut microbiota correlate with many diseases [Manor et al., 2020]
- Many studies conclude that variations in the gut microbiota cause disease
- However, host variables confound disease analyses, i.e. microbiota composition patterns are associated with several host variables [Vujkovic-Cvijin et al., 2020]

Gene Regulatory Networks and Causal Graphs I



Gene Regulatory Networks (GRNs):

- Represent how genes (and their products) regulate each other
- Nodes: genes; directed edges: regulatory influence (activation/repression)
- Often inferred from gene expression data

Gene Regulatory Networks and Causal Graphs II

Connection to Causal Graphs:

- GRNs are inherently causal: they describe mechanisms
- Can be represented as directed acyclic graphs (DAGs)
- Intervening on a gene (e.g., knockout, overexpression) corresponds to using the do-operator

Key Point: GRNs are not just correlation networks (e.g. Bayesian networks) — they aim to capture the true *causal structure* of gene regulation

Example Intervention

Suppose we intervene on gene G_1 to externally increase its expression level. This corresponds to the intervention $(G_1 = \text{high})$.

- Since $G_1 \to G_2$, increasing G_1 will activate G_2
- G_3 represses G_2, but this does not change directly under the intervention unless G_3 is also affected
- As $G_2 \to G_4$, we expect G_4's expression to increase indirectly due to the activation cascade
- The causal effect of G_1 on G_4 can thus be estimated via the pathway $G_1 \to G_2 \to G_4$

This type of reasoning is formalized using the do-operator framework: we aim to estimate $\mathrm{pr}(G_4 \mid \mathrm{do}(G_1 = \text{high}))$

Statistical Symmetry, Causal Asymmetry I

Statistical model: Given joint data on e.g. diet quality (X) and healthiness (Y), we can fit either direction:

$$Y = aX + \varepsilon_Y \quad \text{(forward model)}$$
$$X = bY + \varepsilon_X \quad \text{(reverse or inverse model)}$$

- Both directions can be estimated using linear regression
- Symmetry: No difference between cause and effect in the model
- However, in more complex applications, the forward model is typically much easier to define, whereas the inverse model suffers from information loss (i.e. Y typically contains much less information than X)

Statistical Symmetry, Causal Asymmetry II

Causal model: Direction matters!

- Only the forward model $X \to Y$ tells us what happens to Y if we interwene on X, i.e. when we actively set the value of X in an experiment
- Reversing the direction (i.e., using $Y \to X$) does not yield valid counterfactuals or effects of actions

Key point: Statistical models treat X and Y symmetrically - causal models do not

The Ladder of Causation I

The ladder of causation [Pearl and Mackenzie, 2018]:

- **1. Association (Seeing)**
  - In observational studies, we often begin by measuring the association between treatment and outcome
  - However, association does not imply causation; confounders may be influencing both treatment and outcome
  - Example Question: "Is there an association between taking a new drug and faster recovery from illness?"
- **2. Intervention (Doing)**
  - The "do" operator represents an intervention where we force the treatment, independent of confounders

The Ladder of Causation II

- In randomized experiments, the "do" operator models the action of assigning treatment to subjects, eliminating confounding factors
- This allows for estimating causal effects directly by comparing the treated and untreated groups
- Example Question: "What is the effect of the new drug on recovery time if we intervene and assign the drug to some individuals while others do not receive it?"

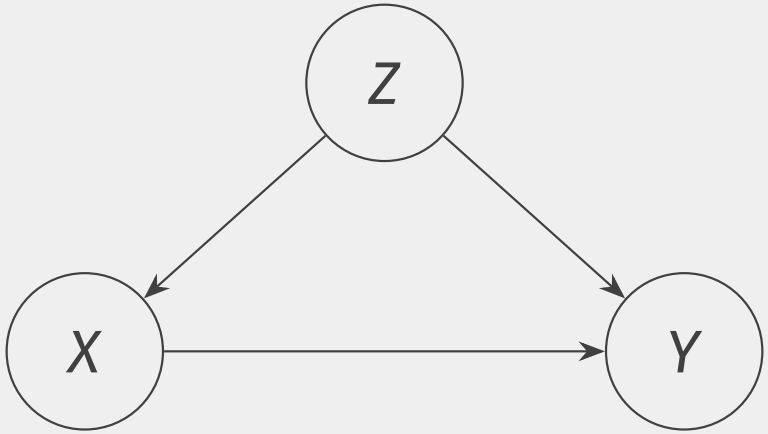
- **3. Counterfactuals (Imagining)**
  - Counterfactuals represent the potential outcomes in a world where the treatment is different

The Ladder of Causation III

- Counterfactuals are generally concerned with individual outcomes, specifically, what would have happened to an individual under a different treatment or intervention scenario
- Example Question: "What would the recovery time have been for a patient who took the new drug if they hadn’t received it?"

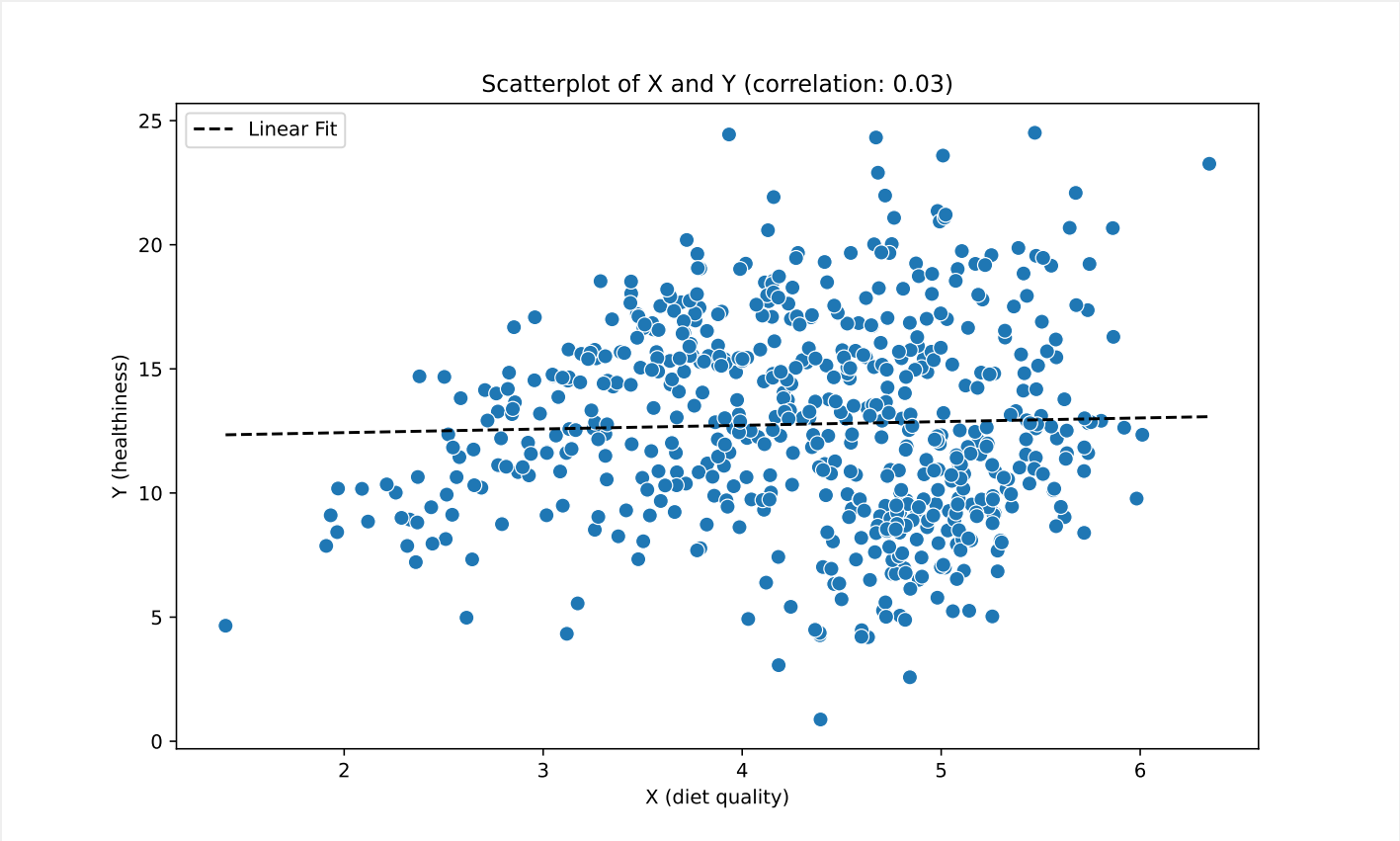
Causal Graphs

Confounder

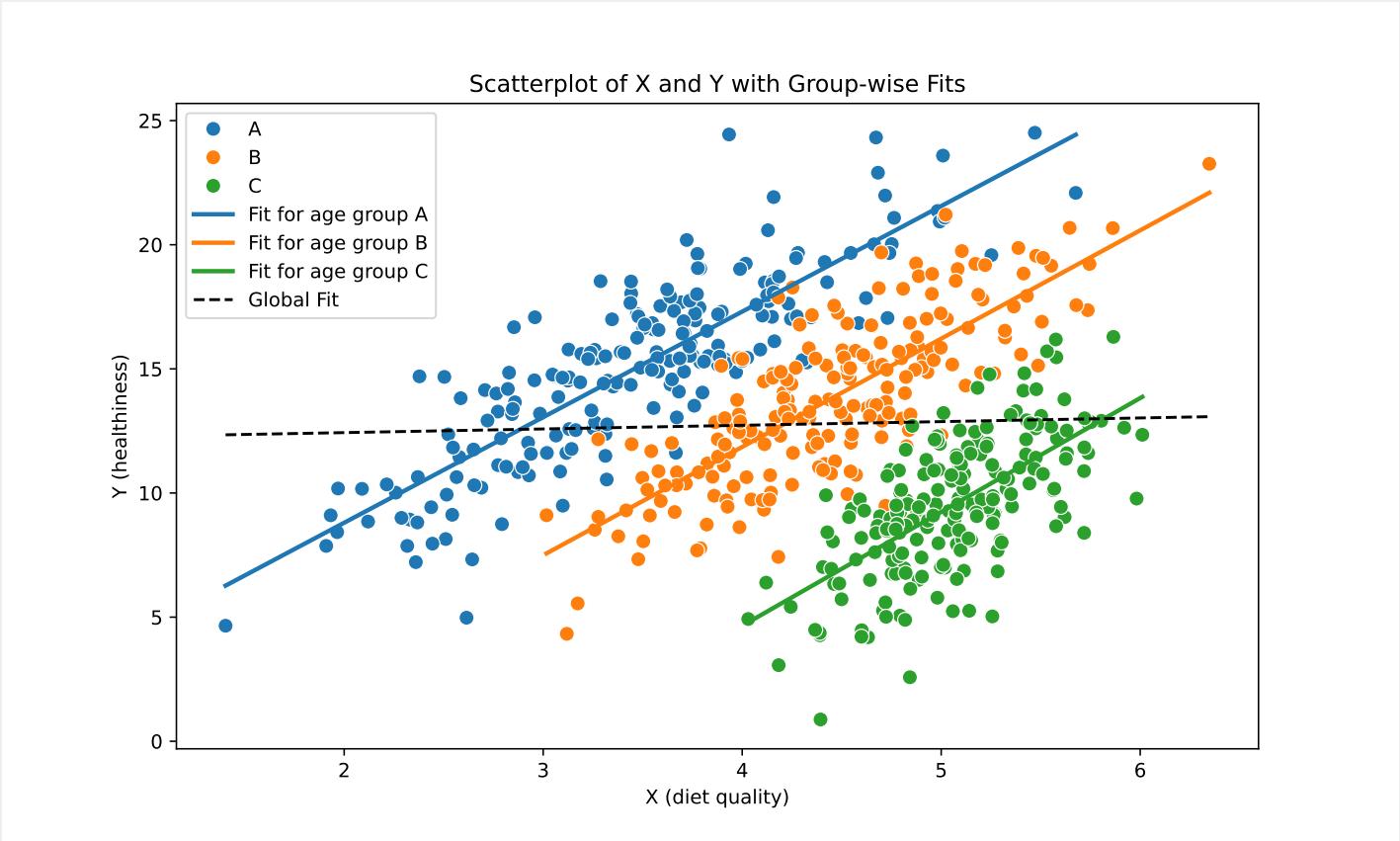


- A confounder (Z) is a variable that influences both the dependent (Y) and independent variable (X)
- A confounder can shadow a causal relationship between the two variables (X, Y)
- It (Z) can create a false impression of a relationship between the two variables (X, Y)

Confounder - Simpson's Paradox



Confounder - Simpson's Paradox



Confounder - Simpson's Paradox - Example

- In a study [Bickel et al., 1977] the authors analyzed 1973 graduate admissions data at UC Berkeley and found that while men appeared to be admitted at a higher rate
- However, department-level data showed little or no bias - and in some cases, a bias in favor of women
- The paradox arose because women tended to apply to more competitive departments with lower admission rates overall

Do-operator I

- The do-operator $\mathrm{do}(X = x)$ simulates an intervention, it sets a variable, rather than just observing it
- Instead of asking:

  "What is the probability of being admitted given gender?"

  we ask:

  "What is the probability of being admitted if we force the applicant to be male or female?"
- Formally:

$$\mathrm{pr}(\mathit{Admitted} \mid \mathrm{do}(\mathit{Gender} = \mathit{female}))$$

Do-operator II

- This removes the backdoor path from

$$\textit{Gender} \to \textit{Department} \to \textit{Admission}$$

  that confounds the correlation.

- To estimate the causal effect of gender on admission, you adjust for the confounding variable (Department). Using backdoor adjustment:

$$\mathrm{pr}(\textit{Admitted} \mid \mathrm{do}(\textit{Gender})) = \sum_{\textit{Dept}} \mathrm{pr}(\textit{Admitted} \mid \textit{Gender}, \textit{Dept}) \cdot \mathrm{pr}(\textit{Dept})$$

- Notice we weight by $\mathrm{pr}(\textit{Dept})$, and not $\mathrm{pr}(\textit{Dept} \mid \textit{Gender})$, which is the key: we simulate the intervention while neutralizing the confounder

Formal View with the do-Operator I

- In observational studies, we can estimate:

$$\mathrm{pr}(Y \mid X = x)$$

  which reflects the association between X and Y. However, this quantity may be confounded by other variables that affect both X and Y

- In contrast, what we really want for causal inference is:

$$\mathrm{pr}(Y \mid \mathrm{do}(X = x))$$

  which represents the distribution of Y under an intervention that sets X to x, breaking any causal paths into X

Formal View with the do-Operator II

- In a randomized controlled trial, treatment X is assigned randomly and independently of all other variables. This effectively removes all incoming edges into X in the causal graph and *determines the distribution of* X. As a result, the distribution of X is independent of potential confounders, and we have:

$$\mathrm{pr}(Y \mid X = x) = \mathrm{pr}(Y \mid \mathrm{do}(X = x))$$

because randomization guarantees that X is not confounded

Confounding - Definition

Confounding

Confounding occurs when:

$$\mathrm{pr}(Y \mid X) \neq \mathrm{pr}(Y \mid \mathrm{do}(X))$$

That is, the observed association between X and Y does not reflect the true causal effect of X on Y, due to the influence of other variables

Backdoor Adjustment - Definition

Backdoor Adjustment

To estimate the causal effect of a variable X on an outcome Y, we can use the backdoor adjustment formula if there exists a *known* set of observed variables Z such that:

$$\mathrm{pr}(Y \mid \mathrm{do}(X)) = \sum_{z} \mathrm{pr}(Y \mid X, Z = z)\,\mathrm{pr}(Z = z)$$

This equality holds if and only if:

- The set Z blocks all backdoor paths from X to Y in the causal graph
- There are no unmeasured confounders affecting both X and Y that are not included in Z

Mediator I

- A **mediator** lies *on the causal path* from X to Y:

$$X \to M \to Y$$

- If we adjust for M, we block part of the causal effect from X to Y
- This leads to a biased estimate of the **total effect**

Mediator II

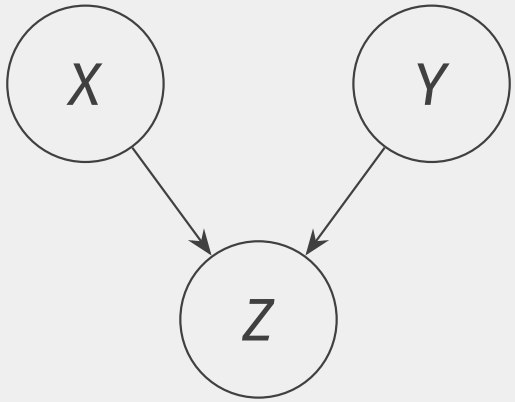
Key Point

Adjusting for a mediator is appropriate only when estimating the **direct effect** of X on Y. For the **total effect**, do *not* adjust for mediators

Example:

- X: Taking a drug
- M: Blood pressure reduction
- Y: Risk of heart attack
- Adjusting for M hides part of the benefit of the drug

Collider I



- A **collider** is a variable with two incoming arrows:

$$X \rightarrow Z \leftarrow Y$$

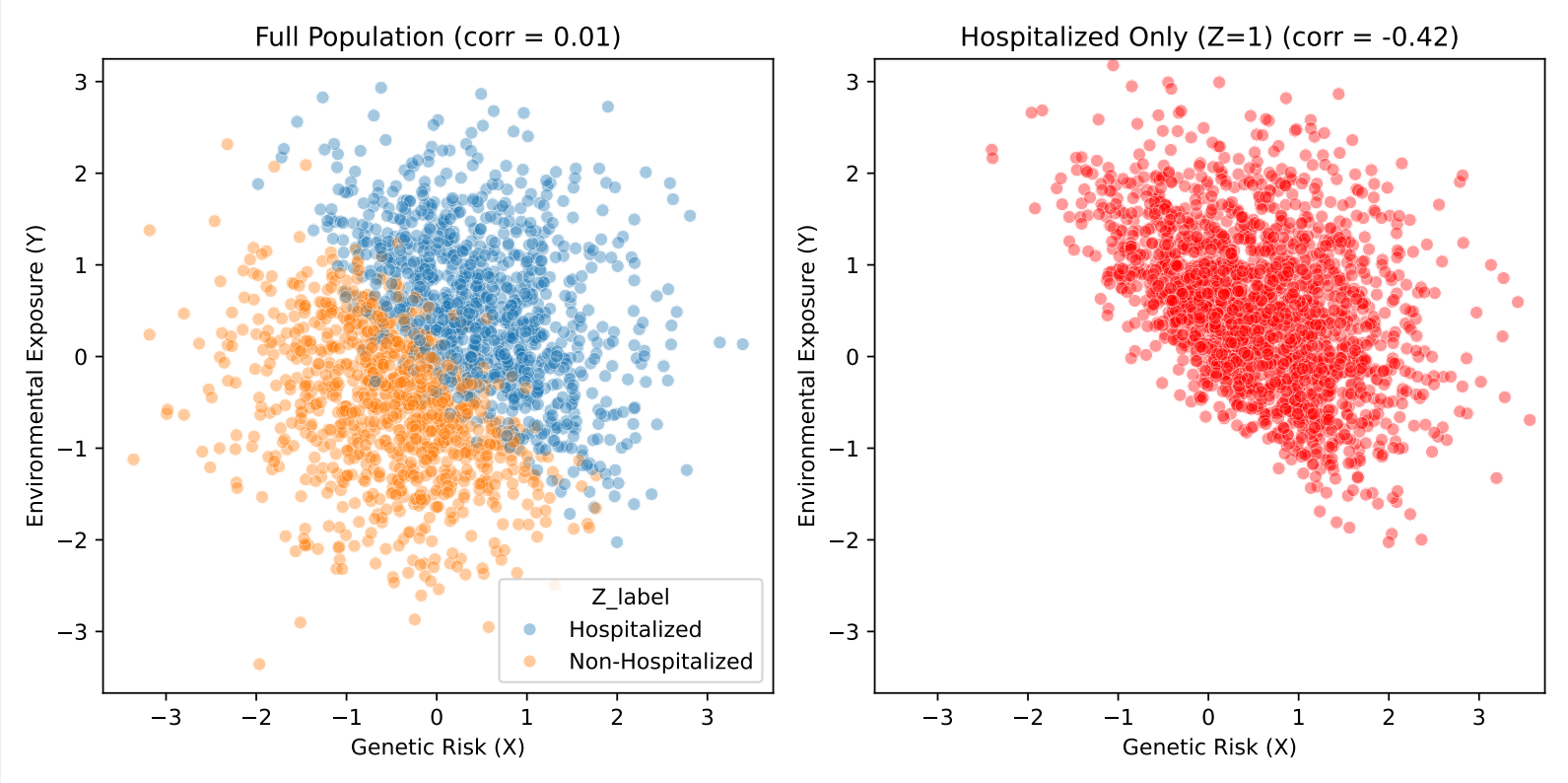
- Without conditioning, the path from X to Y is **blocked**
- Conditioning on Z opens the path, creating a **spurious association**

Collider II

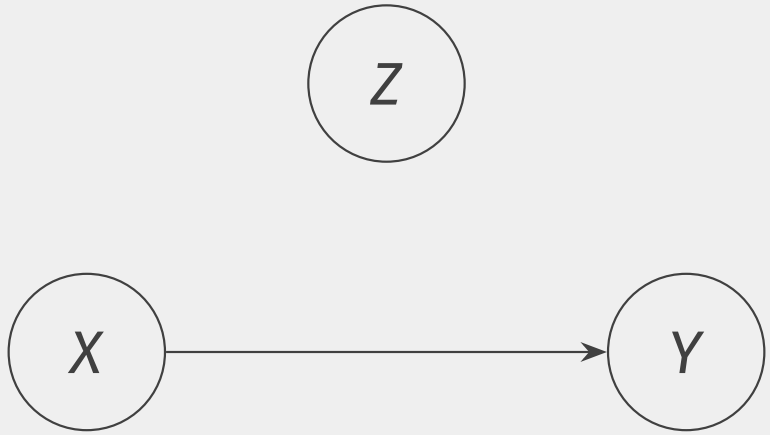
Real-World Example: Collider Bias

- X: Genetic risk of disease
- Y: Environmental exposure
- Z: Hospitalization, caused by both X and Y
- In the general population, X and Y are independent
- But if we only analyze hospitalized patients ($Z = 1$), we may observe a spurious negative correlation between X and Y

Collider III

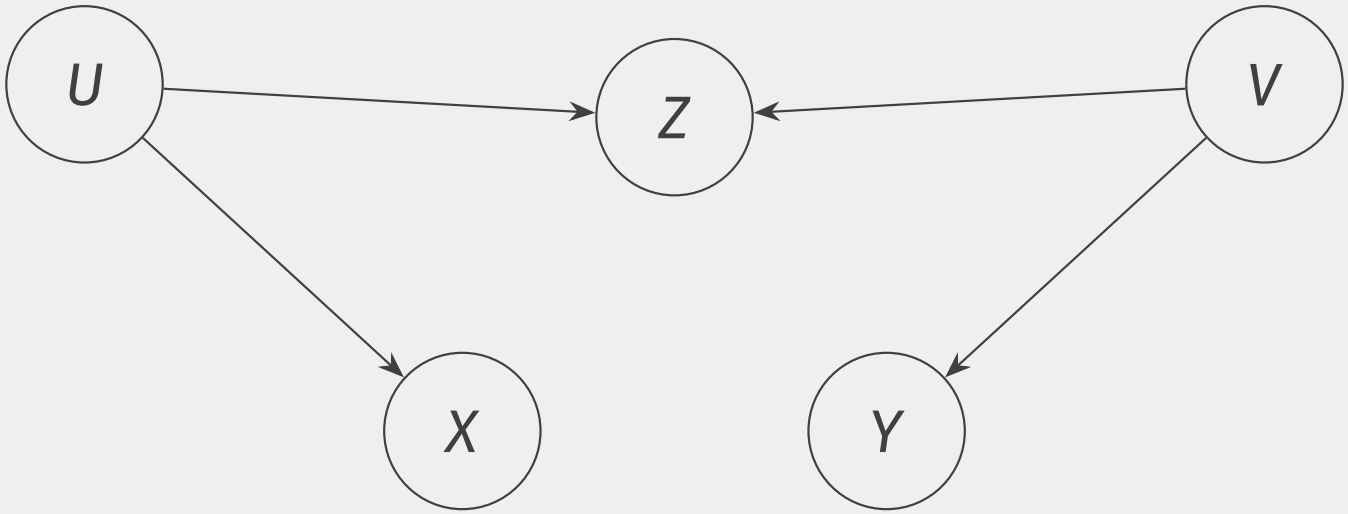


Bystander



- The variable Z is called a bystander and adjustment for Z is not required
- The estimated causal effect $\mathrm{pr}(Y \mid \mathrm{do}(X))$ and adjusting for Z remains valid, but the estimate might be noisier (i.e., higher variance) or harder to interpret due to added model complexity

Bystander



- Z is a **collider** on the path $X \leftarrow U \rightarrow Z \leftarrow V \rightarrow Y$
- Normally, this path is blocked because of the collider structure
- But adjusting for Z **opens** the path, introducing spurious association between X and Y
- So, **adjusting for Z biases the estimate of the causal effect of X on Y**

Backdoor Paths I

- A **backdoor path** is any path from X to Y that starts with an arrow **into** X
- These paths can create **spurious associations** that bias causal estimates
- To estimate the causal effect of X on Y, we must **block all backdoor paths**
- This is done by **adjusting for appropriate variables** - typically confounders

Backdoor Paths II

The backdoor criterion tells us which set of variables to adjust for so that all backdoor paths from X to Y are blocked

Backdoor Criterion [Pearl, 2009]

A set of variables Z satisfies the backdoor criterion if:

- No node in Z is a descendant of X, and
- Z **blocks** all backdoor paths from X to Y

Backdoor Paths III

Blocking [Pearl, 2009]

All paths from X to Y are **blocked** by a conditioning set Z if at least one of the following holds for each path:

- There is a non-collider on the path ($\rightarrow M \rightarrow$ or $\leftarrow M \rightarrow$) such that $M \in Z$
- There is a collider ($\rightarrow C \leftarrow$) such that neither C nor any of its descendants are in Z

Recall: Given a DAG with variables X, Y, Z, we say that X and Y are d**-separated** by Z if all paths between X and Y are blocked when conditioning on Z.

Estimating Causal Effects

Average Treatment Effect (ATE) I

- The **Average Treatment Effect (ATE)** quantifies the expected change in the outcome Y if we were to intervene and set the treatment variable X to different values
- We assume that $X \in \{0, 1\}$ is a **binary treatment** variable
- Formally:

$$\mathrm{ATE} = \mathbb{E}[Y \mid \mathrm{do}(X = 1)] - \mathbb{E}[Y \mid \mathrm{do}(X = 0)]$$

- The ATE represents a causal effect, not merely an association, and requires assumptions to be identified from observational data

Average Treatment Effect (ATE) II

Backdoor Adjustment (Discrete Case)

If we observe a set of covariates Z that satisfies the **backdoor criterion**, the ATE can be estimated as:

$$\mathbb{E}[Y \mid \text{do}(X)] = \sum_{z} \mathbb{E}[Y \mid X, Z = z] \cdot \text{pr}(Z = z)$$

Assuming Z is discrete; for continuous Z, replace the sum with an integral

- Machine learning models can be used to model $X, Z \rightarrow Y$ and to compute $\mathbb{E}[Y \mid X, Z]$

Double Machine Learning I

Goal: Estimate the causal effect of treatment X on outcome Y, adjusting for confounders Z, using flexible machine learning models.

Double machine learning (DML) [Chernozhukov et al., 2018] removes the part of the treatment and outcome that is predictable from confounders, so that the remaining variation is "clean" and uncorrelated with bias (i.e. it orthogonalizes the causal effect estimation):

1. Estimating the nuisance components:
   - $\hat{m}(Z) \approx \mathrm{pr}(X \mid Z)$ — treatment model (propensity)
   - $\hat{g}(Z) \approx \mathrm{pr}(Y \mid Z)$ — outcome model

Double Machine Learning II

2. Residualizing:

$$\tilde{X} = X - \hat{m}(Z), \quad \tilde{Y} = Y - \hat{g}(Z)$$

3. Estimating the causal effect:

$$\tilde{Y} = \theta \cdot \tilde{X} + \varepsilon$$

Why it works

- Removes confounding bias via ML-adjustment
- Final linear step ensures valid inference for θ
- Cross-fitting avoids overfitting and preserves consistency

Further Methods and Reading I

- **Causal Forests:**
  - Extension of random forests for causal inference [Wager and Athey, 2018, Athey and Wager, 2019]
  - Estimates heterogeneous treatment effects (HTEs) by partitioning the data into subgroups with similar treatment responses

- **Bayesian Additive Regression Trees (BART):**
  - BART [Hill, 2011] simplifies model fitting, handles a large number of predictors, and models continuous treatment variables effectively.
  - Provides coherent uncertainty intervals and handles missing data in the outcome variable

Further Methods and Reading II

- **Books:**
  - Causality [Pearl, 2009]
  - Causality for Machine Learning [Schölkopf, 2022]

The End

"Statistics are the poetry of science."

F. Emerson Andrews

References I

Athey, S. and Wager, S. (2019).
Estimating treatment effects with causal forests: An application.
Observational studies, 5(2):37–51.

Bareinboim, E. (2025).
Causal Artificial Intelligence: A Roadmap for Building Causally Intelligent Systems.
https://causalai-book.net/.

Bickel, P. J., Hammel, E. A., and O'Connell, J. W. (1977).
Sex bias in graduate admissions: Data from berkeley.
Statistics and public policy, pages 113–130.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018).
Double/debiased machine learning for treatment and structural parameters.
The Econometrics Journal, 21(1):C1–C68.

References II


Hill, J. L. (2011).
Bayesian nonparametric modeling for causal inference.
Journal of Computational and Graphical Statistics, 20(1):217–240.


Manor, O., Dai, C. L., Kornilov, S. A., Smith, B., Price, N. D., Lovejoy, J. C., Gibbons, S. M., and Magis, A. T. (2020).
Health and disease markers correlate with gut microbiome composition across thousands of people.
Nature communications, 11(1):5206.


Pearl, J. (2009).
***Causality*.**
Cambridge university press.


Pearl, J. and Mackenzie, D. (2018).
***The book of why: the new science of cause and effect*.**
Basic books.

References III


Schölkopf, B. (2022).
Causality for machine learning.
In *Probabilistic and causal inference: The works of Judea Pearl*, pages 765–804.


Vujkovic-Cvijin, I., Sklar, J., Jiang, L., Natarajan, L., Knight, R., and Belkaid, Y. (2020).
Host variables confound gut microbiota studies of human disease.
Nature, 587(7834):448–454.


Wager, S. and Athey, S. (2018).
Estimation and inference of heterogeneous treatment effects using random forests.
Journal of the American Statistical Association, 113(523):1228–1242.