Sensitivity Analysis of Linear Structural Causal Models

Carlos Cinelli
UCLA

Joint work with Daniel Kumor, Bryant Chen, Judea Pearl and Elias Bareinboim

ICML, Long Beach, June 2019

Motivating example: smoking and cancer

Let’s start with a motivating example: the debate
on cigarette smoking and lung cancer (50’s/60’s).

Motivating example: smoking and cancer

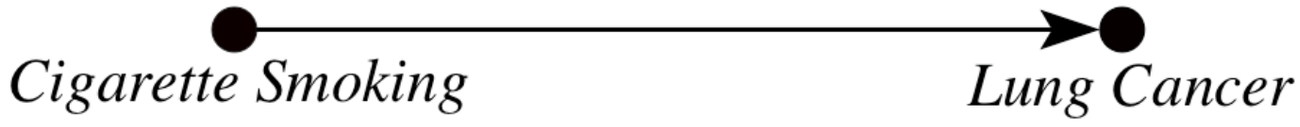
Let's start with a motivating example: the debate on cigarette smoking and lung cancer (50's/60's).

Strong association: smokers had 9 times the risk of nonsmokers to develop lung cancer.

Motivating example: smoking and cancer

Let's start with a motivating example: the debate on cigarette smoking and lung cancer (50's/60's).

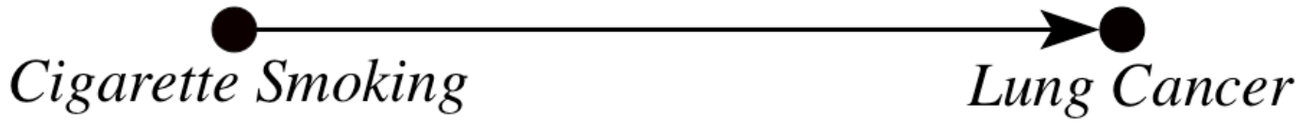
Strong association: smokers had 9 times the risk of nonsmokers to develop lung cancer. Causal?



Motivating example: smoking and cancer

Let's start with a motivating example: the debate on cigarette smoking and lung cancer (50's/60's).

Strong association: smokers had 9 times the risk of nonsmokers to develop lung cancer. Causal?

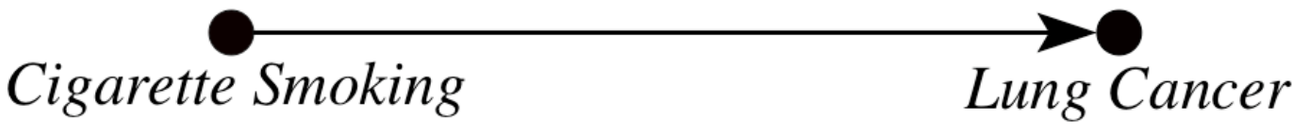


Not everyone agreed with this hypothesis.

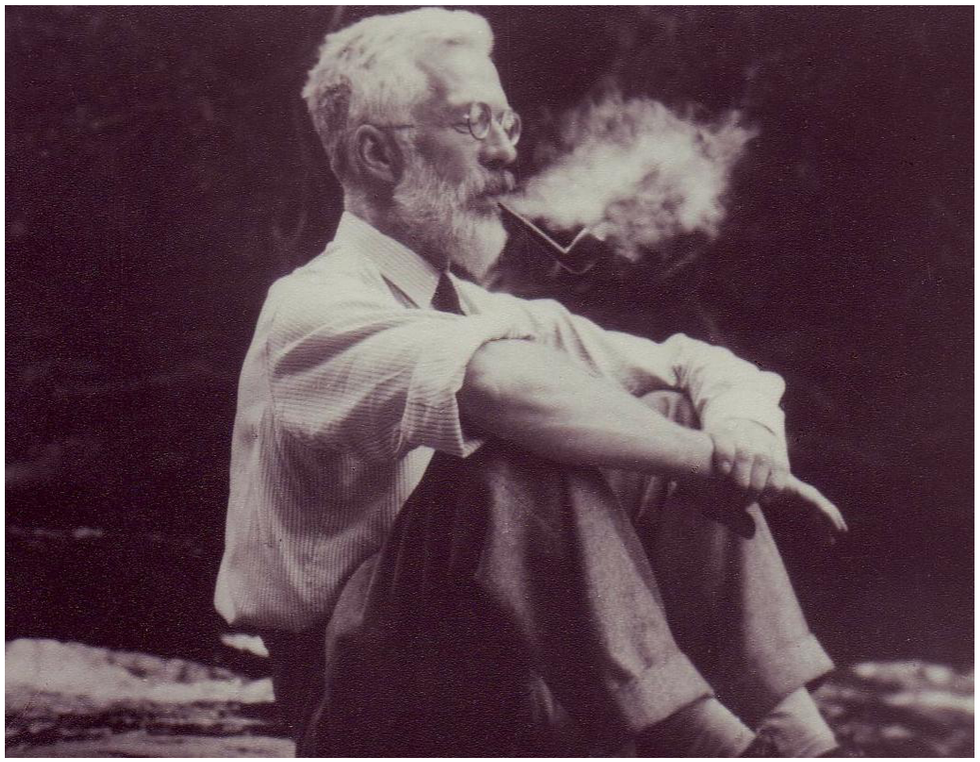
Motivating example: smoking and cancer

Let's start with a motivating example: the debate on cigarette smoking and lung cancer (50's/60's).

Strong association: smokers had 9 times the risk of nonsmokers to develop lung cancer. Causal?



Not everyone agreed with this hypothesis.



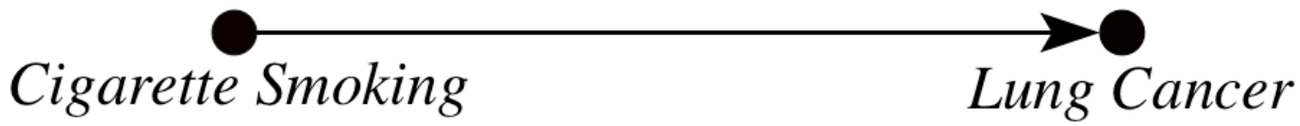
"For my part, I think it is more likely that a ***common cause*** supplies the explanation… The obvious common cause to think of is the ***genotype***"

- Ronald Fisher (1958)

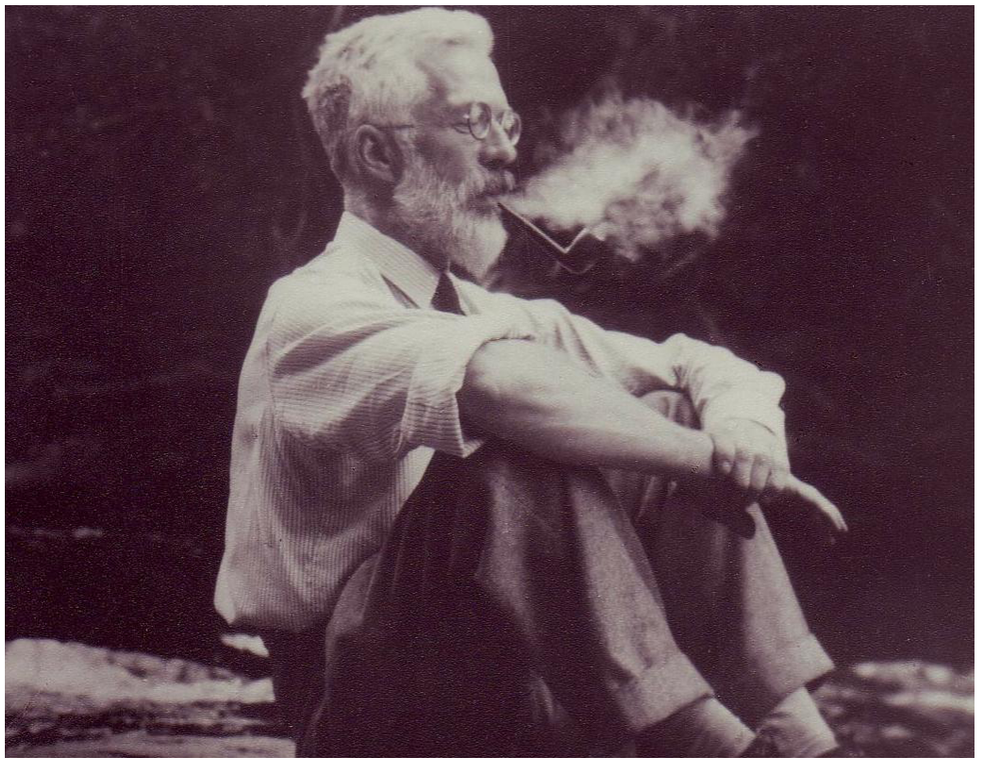
Motivating example: smoking and cancer

Let's start with a motivating example: the debate on cigarette smoking and lung cancer (50's/60's).

Strong association: smokers had 9 times the risk of nonsmokers to develop lung cancer. Causal?



Not everyone agreed with this hypothesis.



"For my part, I think it is more likely that a ***common cause*** supplies the explanation… The obvious common cause to think of is the ***genotype***"

- Ronald Fisher (1958)

Observational data alone cannot distinguish both models.

Motivating example: smoking and cancer

Motivating example: smoking and cancer

Let's suppose for a moment that Fisher's hypothesis were true.

Motivating example: smoking and cancer

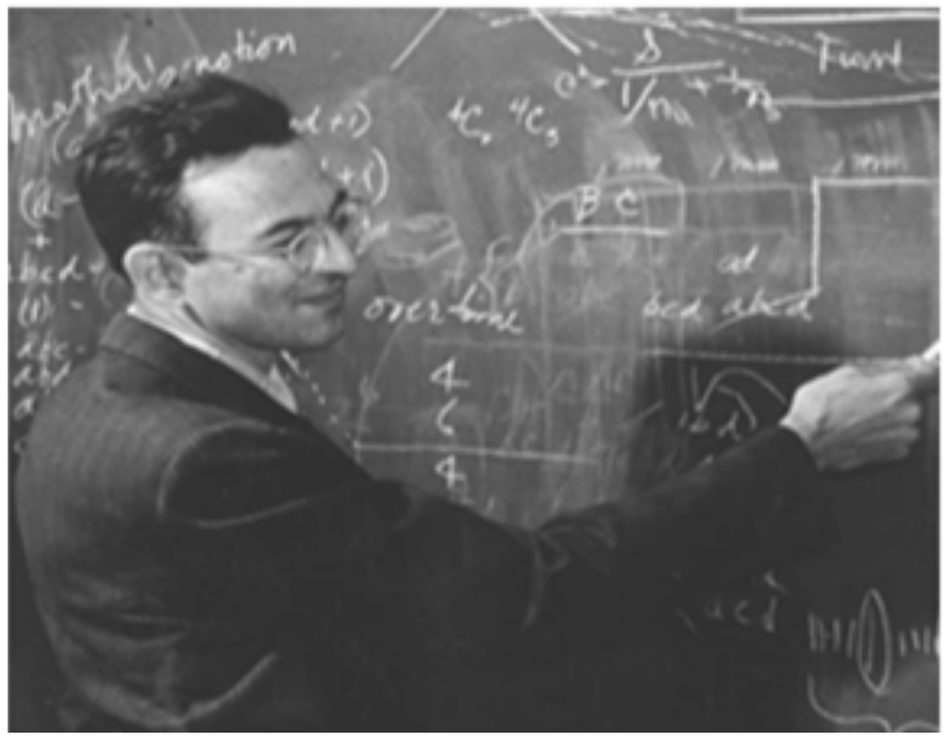
Let's suppose for a moment that Fisher's hypothesis were true.

How strong would unobserved confounding need to be to explain all the observed association?

Motivating example: smoking and cancer

Let's suppose for a moment that Fisher's hypothesis were true.

How strong would unobserved confounding need to be to explain all the observed association?

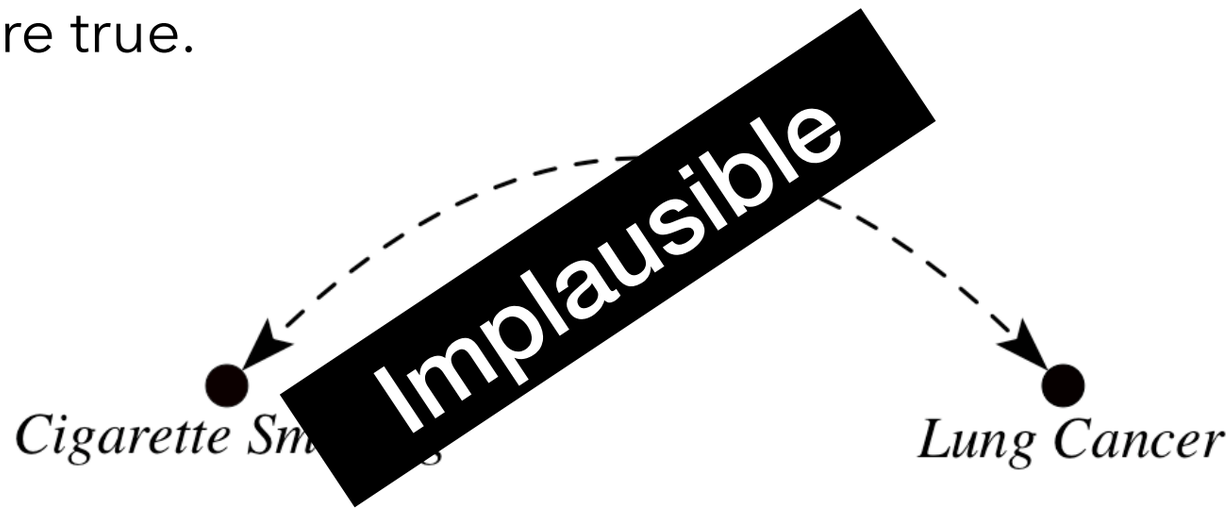


"…if cigarette smokers have 9 times the risk of nonsmokers for developing lung cancer, ***and this is not because cigarette smoke is a causal agent,*** …, then ***the proportion of hormone-X-producers among cigarette smokers must be at least 9 times greater than that of nonsmokers***"

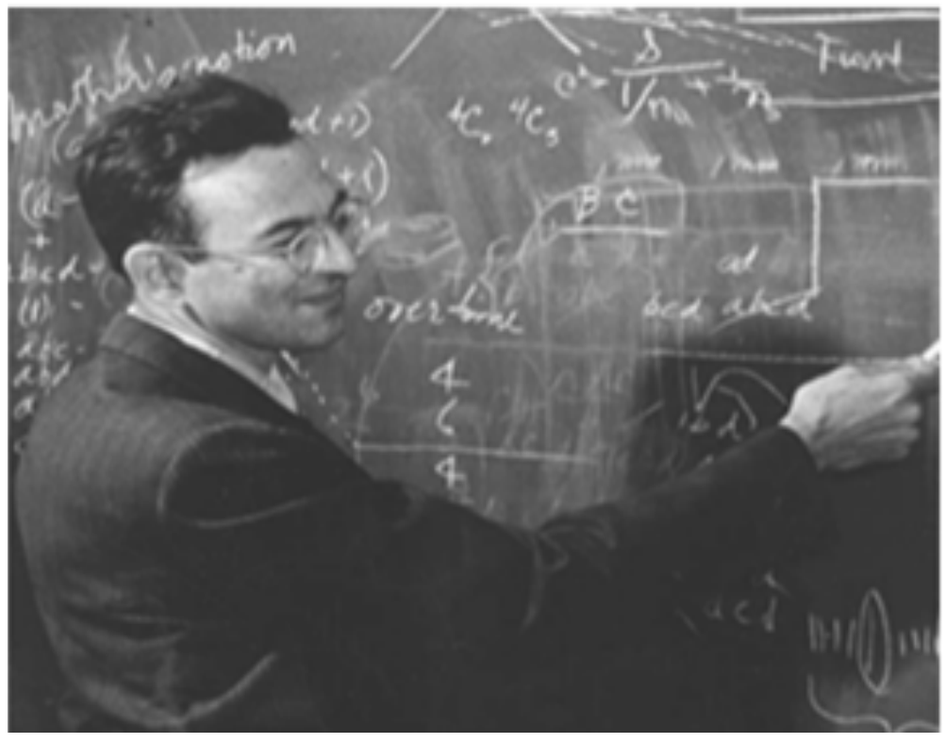
- Cornfield et al (1959)

Motivating example: smoking and cancer

Let's suppose for a moment that Fisher's hypothesis were true.



How strong would unobserved confounding need to be to explain all the observed association?

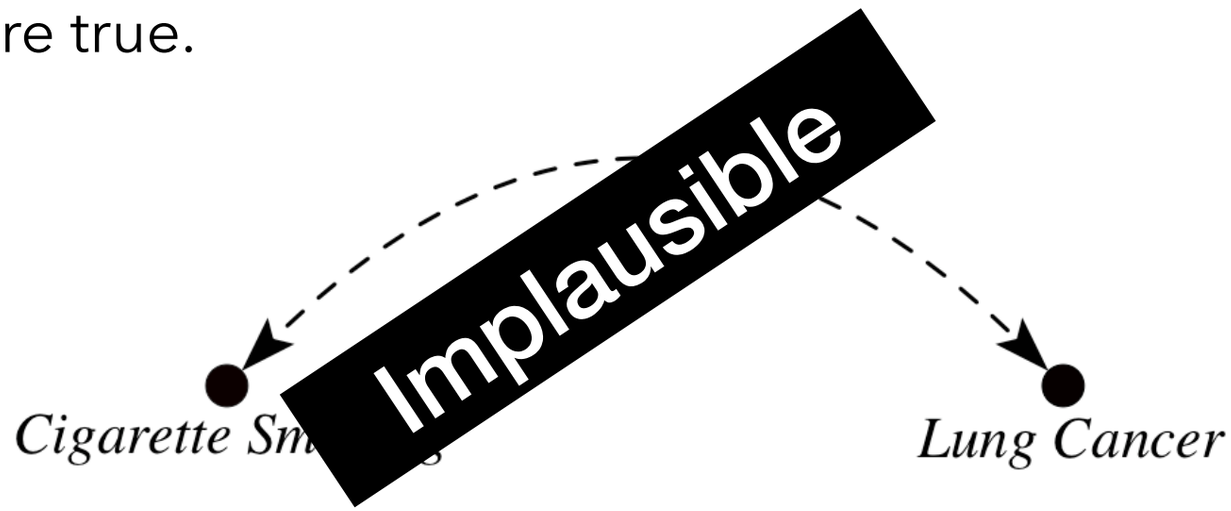


"…if cigarette smokers have 9 times the risk of nonsmokers for developing lung cancer, ***and this is not because cigarette smoke is a causal agent,*** …, then ***the proportion of hormone-X-producers among cigarette smokers must be at least 9 times greater than that of nonsmokers***"

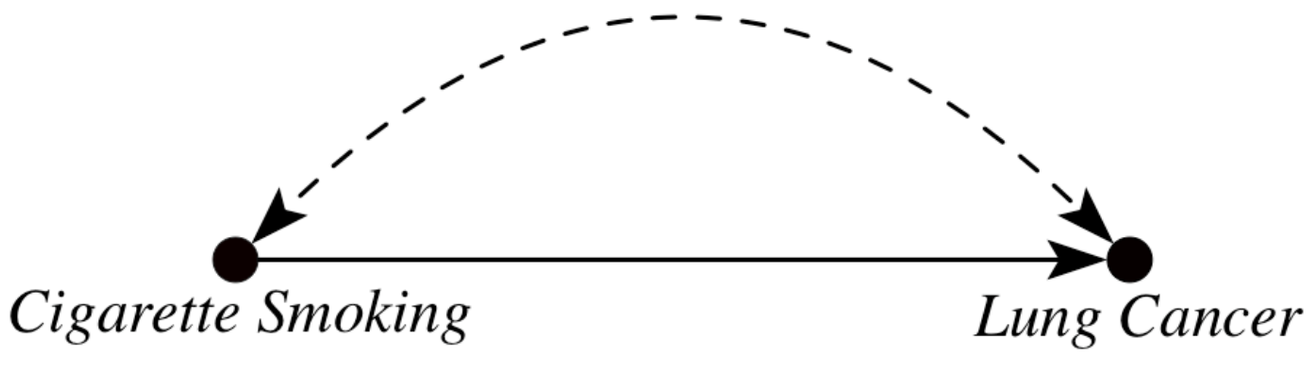
- Cornfield et al (1959)

Motivating example: smoking and cancer

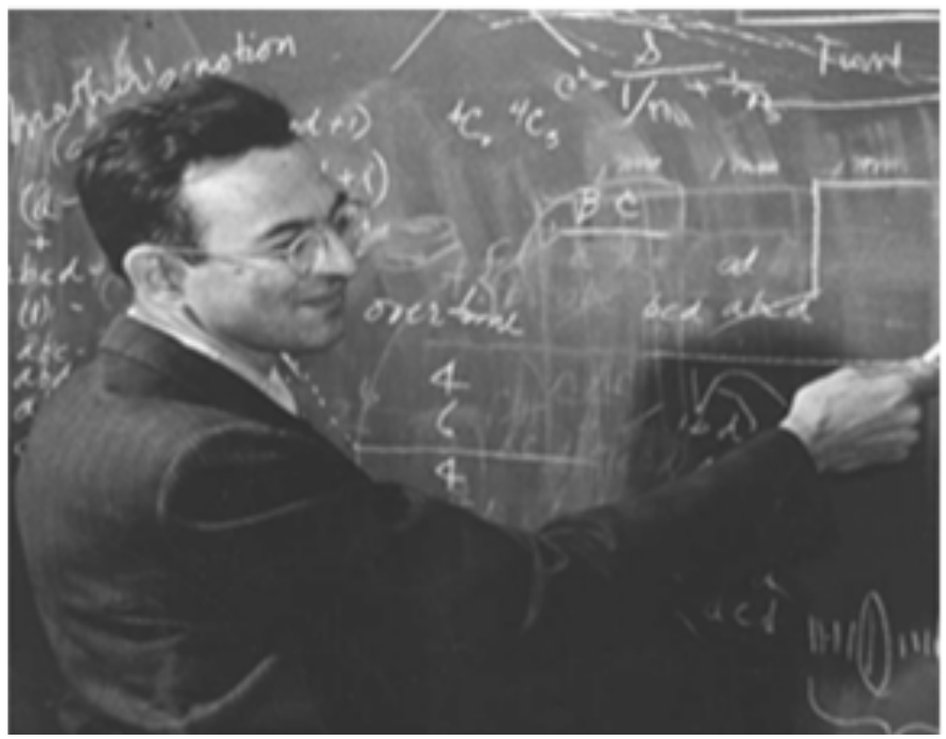
Let's suppose for a moment that Fisher's hypothesis were true.



How strong would unobserved confounding need to be to explain all the observed association?



Sensitivity analysis + plausibility judgments = there must be a causal path between cigarette smoking and lung cancer.



"…if cigarette smokers have 9 times the risk of nonsmokers for developing lung cancer, ***and this is not because cigarette smoke is a causal agent,*** …, then ***the proportion of hormone-X-producers among cigarette smokers must be at least 9 times greater than that of nonsmokers***"

- Cornfield et al (1959)

In summary: why sensitivity analysis?

In summary: why sensitivity analysis?

Causal inference requires (sometimes untestable) ***assumptions about the data generating process***, such as:

(i) the absence of certain direct causal relationships;

(ii) the absence of unobserved common causes between certain variables.

In summary: why sensitivity analysis?

Causal inference requires (sometimes untestable) ***assumptions about the data generating process***, such as:

(i) the absence of certain direct causal relationships;

(ii) the absence of unobserved common causes between certain variables.

Therefore, conclusions based on causal models are ***provisional.*** What if these assumptions are disputed?

In summary: why sensitivity analysis?

Causal inference requires (sometimes untestable) ***assumptions about the data generating process***, such as:

(i) the absence of certain direct causal relationships;

(ii) the absence of unobserved common causes between certain variables.

Therefore, conclusions based on causal models are ***provisional.*** What if these assumptions are disputed?

Sensitivity analysis allows us to ***quantify how the violations of assumptions affect our conclusions.***

In summary: why sensitivity analysis?

Causal inference requires (sometimes untestable) ***assumptions about the data generating process***, such as:

(i) the absence of certain direct causal relationships;

(ii) the absence of unobserved common causes between certain variables.

Therefore, conclusions based on causal models are ***provisional.*** What if these assumptions are disputed?

Sensitivity analysis allows us to ***quantify how the violations of assumptions affect our conclusions.***

These results can then be ***submitted to expert judgment***, to decide whether problematic degrees of violation are plausible.

Current sensitivity analysis literature

Current sensitivity analysis literature

Sensitivity analysis has been extensively studied in the Health Sciences, Economics, Statistics…

Current sensitivity analysis literature

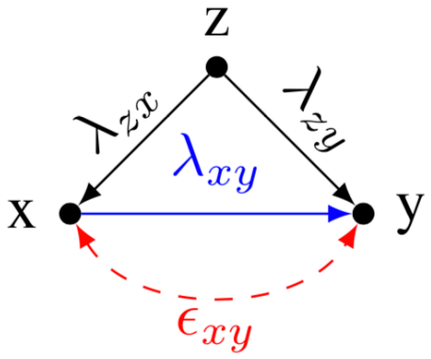
Sensitivity analysis has been extensively studied in the Health Sciences, Economics, Statistics…

However, the current literature is limited to ***specific model structures*** and ***solved on a case-by-case basis;*** e.g., separate results for,

Current sensitivity analysis literature

Sensitivity analysis has been extensively studied in the Health Sciences, Economics, Statistics…

However, the current literature is limited to ***specific model structures*** and ***solved on a case-by-case basis;*** e.g., separate results for,

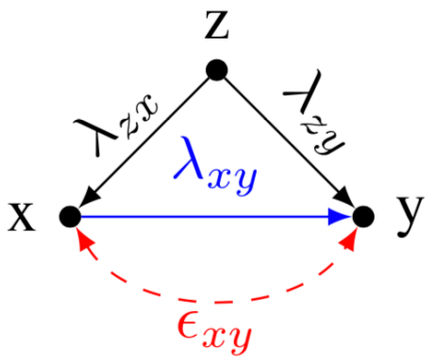


(a) Backdoor

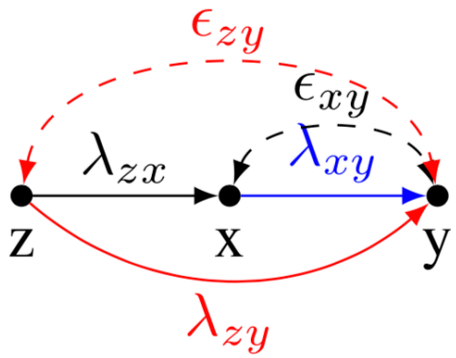
Current sensitivity analysis literature

Sensitivity analysis has been extensively studied in the Health Sciences, Economics, Statistics…

However, the current literature is limited to ***specific model structures*** and ***solved on a case-by-case basis;*** e.g., separate results for,



(a) Backdoor

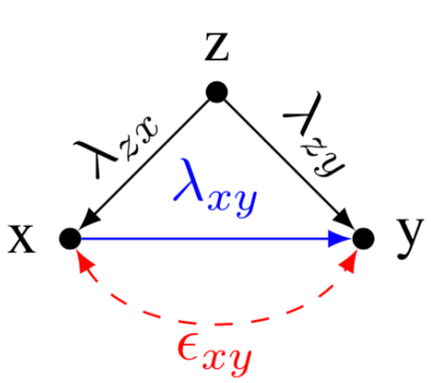


(b) IV

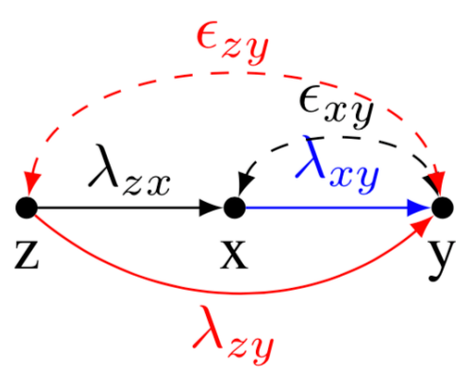
Current sensitivity analysis literature

Sensitivity analysis has been extensively studied in the Health Sciences, Economics, Statistics…

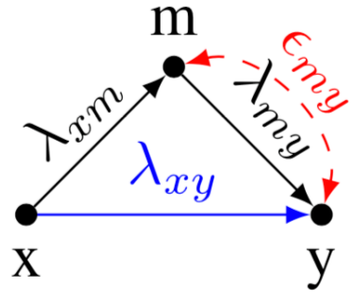
However, the current literature is limited to ***specific model structures*** and ***solved on a case-by-case basis;*** e.g., separate results for,



(a) Backdoor



(b) IV

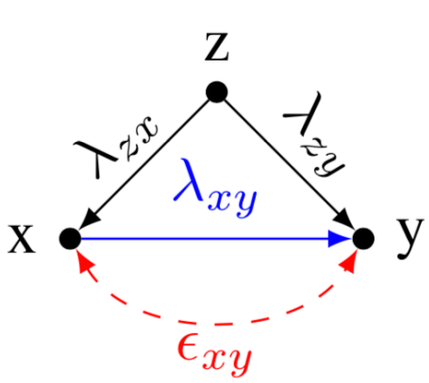


(c) Mediation

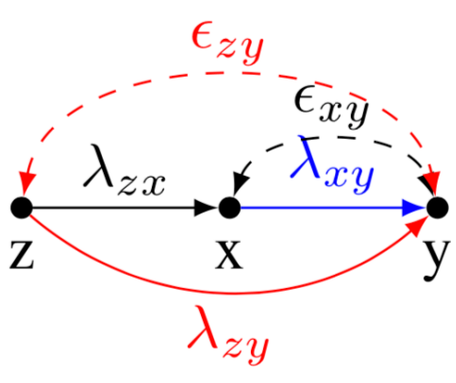
Current sensitivity analysis literature

Sensitivity analysis has been extensively studied in the Health Sciences, Economics, Statistics…

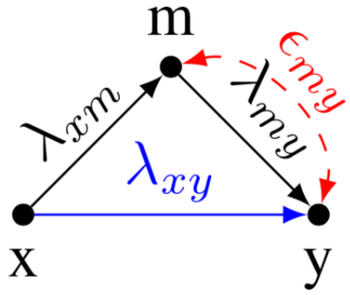
However, the current literature is limited to ***specific model structures*** and ***solved on a case-by-case basis;*** e.g., separate results for,



(a) Backdoor



(b) IV



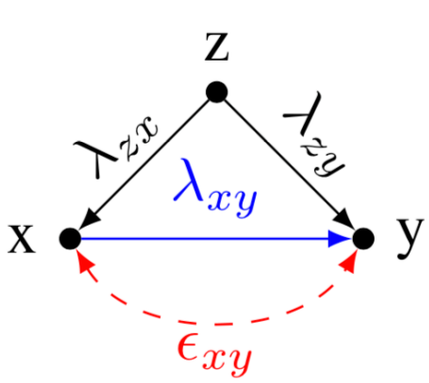
(c) Mediation

And so on.

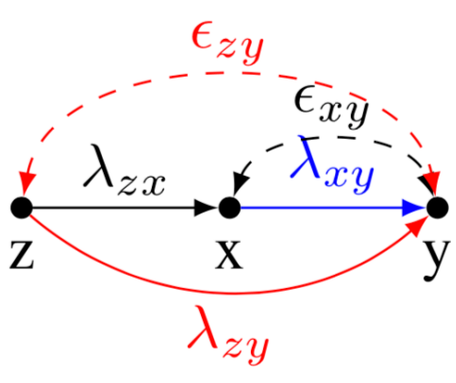
Current sensitivity analysis literature

Sensitivity analysis has been extensively studied in the Health Sciences, Economics, Statistics…

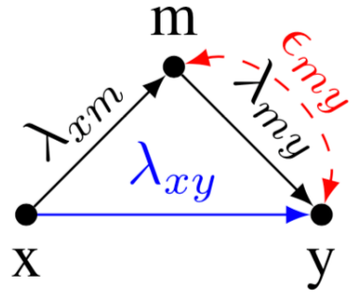
However, the current literature is limited to ***specific model structures*** and ***solved on a case-by-case basis;*** e.g., separate results for,



(a) Backdoor



(b) IV



(c) Mediation

And so on.

Can we have a ***general-purpose, algorithmic framework*** that captures all these canonical cases – and many more?

A systematic approach for sensitivity analysis

Our contribution: *a formal, systematic approach to sensitivity analysis* for *arbitrary linear Structural Causal Models (SCMs).*

A systematic approach for sensitivity analysis

Our contribution: *a formal, systematic approach to sensitivity analysis* for *arbitrary linear Structural Causal Models (SCMs).*

1. Formalize sensitivity analysis as ***identification with non-zero constraints;***

A systematic approach for sensitivity analysis

Our contribution: *a formal, systematic approach to sensitivity analysis* for *arbitrary linear Structural Causal Models (SCMs).*

1. Formalize sensitivity analysis as ***identification with non-zero constraints;***

2. Devise a ***novel graphical procedure (PushForward)*** to incorporate ***numerical constraints on bidirected edges;***

A systematic approach for sensitivity analysis

Our contribution: *a formal, systematic approach to sensitivity analysis* for *arbitrary linear Structural Causal Models (SCMs).*

1. Formalize sensitivity analysis as ***identification with non-zero constraints;***

2. Devise a ***novel graphical procedure (PushForward)*** to incorporate ***numerical constraints on bidirected edges;***

3. Develop an ***efficient graph-based identification algorithm to derive sensitivity curves.***

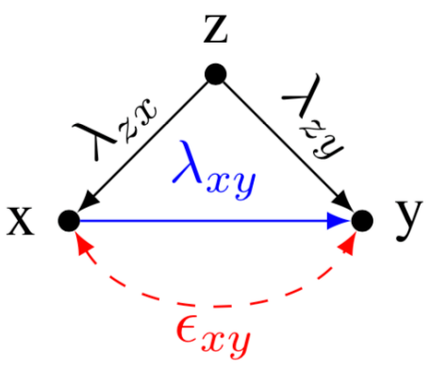
A systematic approach for sensitivity analysis

Our contribution: *a formal, systematic approach to sensitivity analysis* for *arbitrary linear Structural Causal Models (SCMs).*

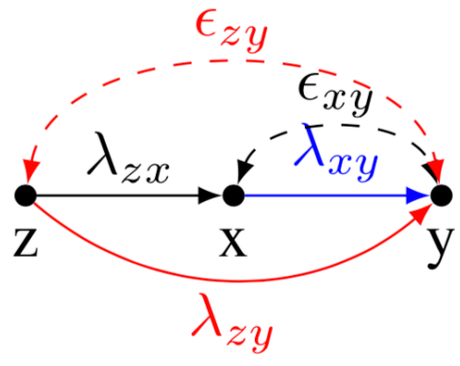
1. Formalize sensitivity analysis as ***identification with non-zero constraints;***

2. Devise a ***novel graphical procedure (PushForward)*** to incorporate ***numerical constraints on bidirected edges;***

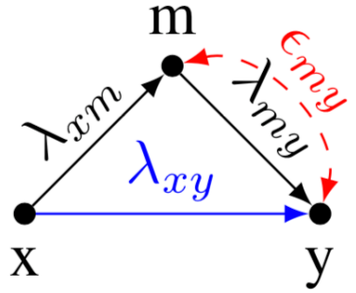
3. Develop an ***efficient graph-based identification algorithm to derive sensitivity curves.***



(a) Backdoor ✓



(b) IV ✓



(c) Mediation ✓

…canonical cases are a small subset of all possible sensitivity analyses covered by our framework.

For details, come to our poster session:

Wed, Jun 12th
6:30pm–9:00pm
@ Pacific Ballroom #78

Or see paper: https://tinyurl.com/y5urlwqs

Thank you!

Contact
carloscinelli@ucla.edu
twitter: @analisereal