



Story of Causality

A Poem composed by Philosophers, Statisticians, and Computer Scientists 🎵

June En-Yu Lai

Institute of Statistical Science, Academia Sinica, Taiwan

April 26, 2022

How do we formulate the idea of “causation”?

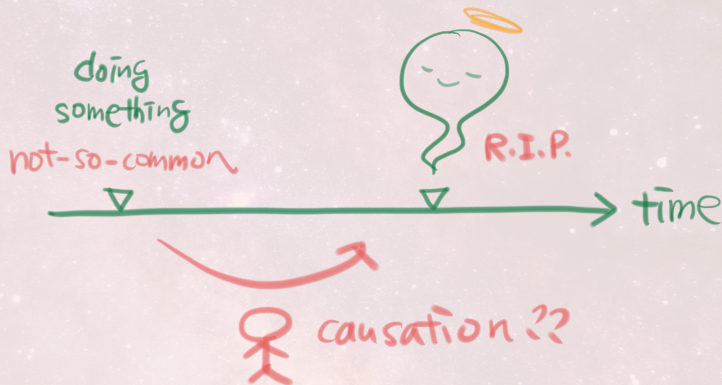
誼誼誼新聞上報導了有人打疫苗後死掉了耶？

- ▶ 時間順序與時程長短
- ▶ 「因」與「果」的『相關性』 (correlation/association)

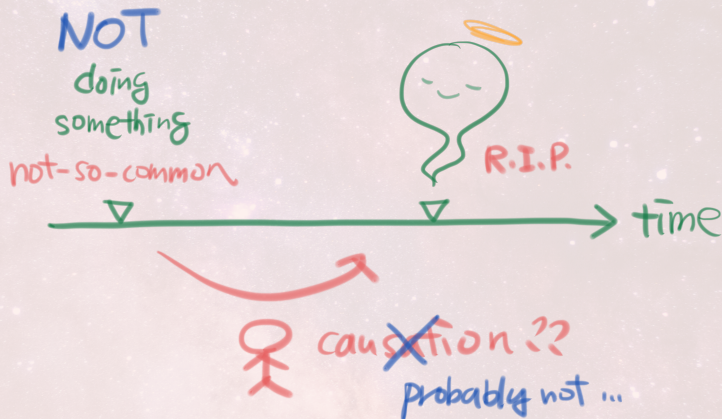
但我聽說考量同年齡層的死亡率，並沒有特別高？

- ▶ 「因」與「果」的『虛擬情況推理』 (counterfactual)

How do we formulate the idea of “causation”?



How do we formulate the idea of “causation”?



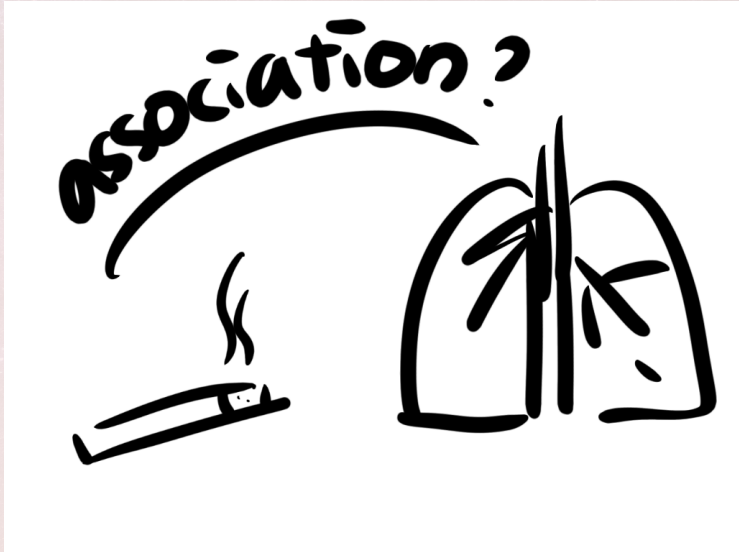
How do we formulate the idea of “causation”?



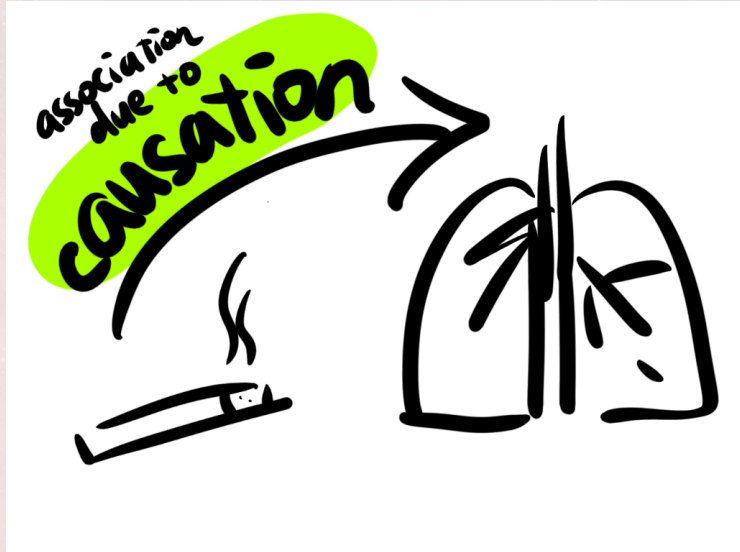
How do we formulate the idea of “causation”?



How do we formulate the idea of “causation”?



How do we formulate the idea of “causation”?



A rough sketch of causation

- ▶ Causation $\rightarrow$ Association
- ▶ Association $\rightarrow$ Causation or Common Cause or mixed structures
- ▶ Association can help us to do prediction!!
- ▶ Causation can help us to make decision!!
- ▶ The idea of **counterfactual** may help us to define causality (as a philosophical stance).

Outline

- ▶ **A Brief History of Causal Models**
 - ▶ Graphical model for association
 - ▶ Potential outcome framework
- ▶ Causal Models in Philosophy
 - ▶ Logic of counterfactuals
 - ▶ Theories of actual causation
- ▶ Causal Models in Machine Learning
 - ▶ Necessary and sufficient conditions
 - ▶ Causal discovery
 - ▶ Counterfactual learning for recommendation system
- ▶ Causal Models in Biostatistics and Epidemiology
 - ▶ Assumptions for causal inference
 - ▶ Mediation analysis

A Brief History of Causal Models

- ▶ Graphical model for association
 - ▶ Sewall Wright
 - computational geneticist
 - path analysis (1921, 1934)
 - ▶ Peter Spirtes, Clark Glymour, and Richard Scheines
 - philosophers (0w0???), CMU
 - causal discovery (Causation, Prediction, and Search, 2000 [SGS])
 - ⇒ +Pitts: **TEGRAD program/causal discovery center, causalMGM (NAR, 2020)**
 - ▶ Judea Pearl
 - computer scientist, UCLA
 - $do(\cdot)$ operator (Causality, 2009)
 - ⇒ axiomatization links to Lewis and Robins
- ▶ Potential outcome framework

Graphical model for association — Sewall Wright

CORRELATION AND CAUSATION

By SEWALL WRIGHT

Senior Animal Husbandman in Animal Genetics, Bureau of Animal Industry, United States Department of Agriculture

PART I. METHOD OF PATH COEFFICIENTS

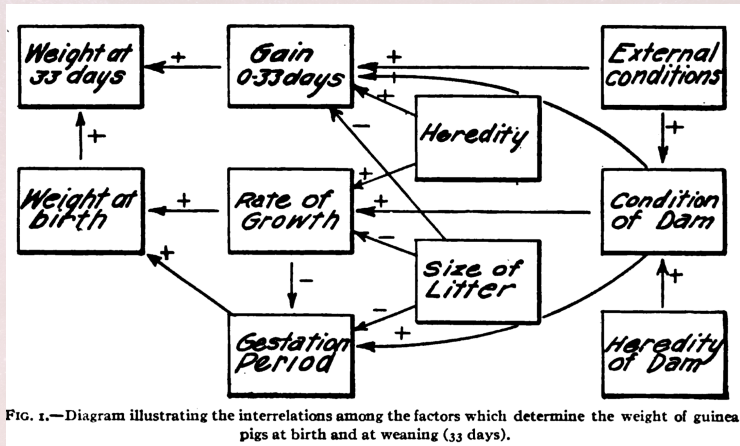
INTRODUCTION

The ideal method of science is the study of the direct influence of one condition on another in experiments in which all other possible causes of variation are eliminated. Unfortunately, causes of variation often seem to be beyond control. In the biological sciences, especially, one often has to deal with a group of characteristics or conditions which are correlated because of a complex of interacting, uncontrollable, and often obscure causes. The degree of correlation between two variables can be calculated by well-known methods, but when it is found it gives merely the resultant of all connecting paths of influence.

The present paper is an attempt to present a method of measuring the direct influence along each separate path in such a system and thus of finding the degree to which variation of a given effect is determined by each particular cause. The method depends on the combination of knowledge of the degrees of correlation among the variables in a system with such knowledge as may be possessed of the causal relations. In cases in which the causal relations are uncertain the method can be used to find the logical consequences of any particular hypothesis in regard to them.

Journal of
Agricultural
Research, 1921

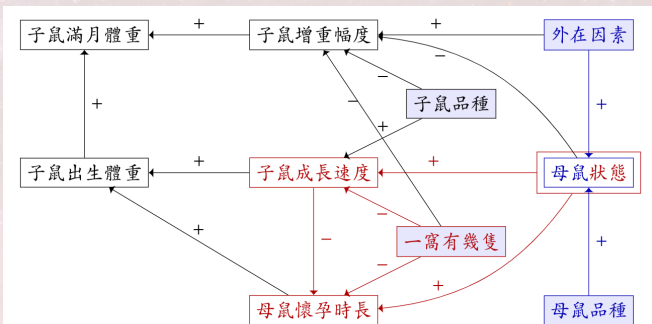
Graphical model for association — Sewall Wright



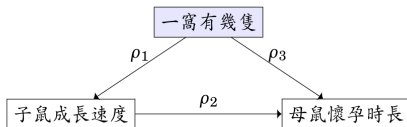
Wright: 豚鼠好可愛！♥♥

Pearl: 這世上無人能阻止 Wright 對數學的愛，如果有，那就是豚鼠！

Graphical model for association — Sewall Wright



(a) 決定子鼠體重的原因及其相互關係的圖形表示法 (Wright, 1921)



(b) 上圖(a)的一個子圖，路徑係數由評估每個邊的關係強度整合而來

A Brief History of Causal Models

- ▶ Graphical model for association
 - ▶ Sewall Wright
 - computational geneticist
 - path analysis (1921, 1934)
 - ▶ Peter Spirtes, Clark Glymour, and Richard Scheines
 - philosophers (0w0???), CMU
 - causal discovery (Causation, Prediction, and Search, 2000 [SGS])
 - ⇒ +Pitts: **TEGRAD program/causal discovery center, causalMGM (NAR, 2020)**
 - ▶ Judea Pearl
 - computer scientist, UCLA
 - $do(\cdot)$ operator (Causality, 2009)
 - ⇒ axiomatization links to Lewis and Robins
- ▶ Potential outcome framework

A Brief History of Causal Models

- ▶ Graphical model for association
- ▶ Potential outcome framework
 - ▶ Jerzy Neyman
 - mathematical statistician
 - potential yield (1923)
 - ▶ Donald Rubin
 - statistician, Harvard
 - causal inference using missing data approaches (1978)
 - ⇒ establish counterfactual notation
 - ▶ James Robins
 - epidemiologist, Harvard
 - causal mediation (1992)
 - ⇒ counterfactual is not a possible world

Statistical Science

1990, Vol. 5, No. 4, 465–480

On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.

Jerzy Splawa-Neyman

Translated and edited by D. M. Dabrowska and T. P. Speed from the Polish original, which appeared in *Roczniki Nauk Rolniczych Tom X (1923) 1–51 (Annals of Agricultural Sciences)*

Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies

Donald B. Rubin

Potential outcome framework — Jerzy Neyman

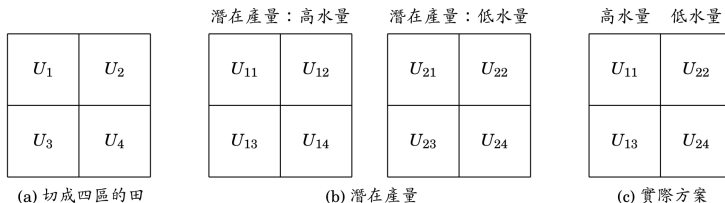


Figure 2: 隨機分配範例。(a)為切成四區的田， U 代表該區的產量，下標為分區的編號。(b)考慮水量高低之下所想像的兩種潛在產量。第一個下標為條件的編號，下標一為高水量，二為低水量。第二個下標為分區的編號。(c)實務上的做法，隨機分派其中兩區使用高水量，另兩區為低水量，並使用這個方案來推估潛在產量的差異。

Outline

- ▶ A Brief History of Causal Models
 - ▶ Graphical model for association
 - ▶ Potential outcome framework
- ▶ **Causal Models in Philosophy**
 - ▶ **Logic of counterfactuals**
 - ▶ Theories of actual causation
- ▶ Causal Models in Machine Learning
 - ▶ Necessary and sufficient conditions
 - ▶ Causal discovery
 - ▶ Counterfactual learning for recommendation system
- ▶ Causal Models in Biostatistics and Epidemiology
 - ▶ Assumptions for causal inference
 - ▶ Mediation analysis

Causal Models in Philosophy — Terminology

- ▶ 個例因果 (token, actual, singular, or individual causation)
「你拉肚子是因為飯前沒洗手」
- ▶ 類型因果 (type, general, population, or average causation)
「抽煙會導致肺癌」
- ▶ 假設條件句 (subjunctive conditional) / Pearl: counterfactual
「假如我之前乖乖洗手的話，就不會拉肚子了」
 - ▶ 也稱為“虛擬條件句”或“反事實條件句” (counterfactual conditional)
 - ▶ 通常該類語句的“前件”會假想某些在現實世界“沒發生”的事件，而“後件”則描述了在該假想事件發生後，可能的結果為何。
- ▶ 指示條件句 (indicative conditional) / Pearl: intervention
「如果我現在去洗手，就不會因此拉肚子」
 - ▶ 指示條件句的前件會假設某些在現實世界“還未發生”的事件，而其後件則描述了在該假想事件發生後，可能的結果為何。

Causal Models in Philosophy

- ▶ 簡介因果模型歷史時，我們提到兩個學派，一是由相關性發展出的圖形模型學派；二是由隨機分配發展出的潛在結果模型。十分有趣的，在因果關係的哲學理論裡也有對應的思維。
- ▶ Hume 定義因果關係時說：“我們或許可以定義『因』為某一事物，它被另一事物所追隨；而所有與前者類似的事物，都被與後者相似的事物追隨。換言之，假如前者不曾存在，後者就根本不會存在。
- ▶ Lewis 在評論這段定義時就說，Hume 所謂的“換言之”根本就是另一個定義：前後兩句話其實就對應到因果關係的“規律性分析” (regularity analysis) 與“虛擬情況分析” (counterfactual analysis)。
- ▶ 可以看出，規律性的定義十分類似於統計學上的相關性，而虛擬情況的定義其實就是潛在結果模型想計算的目標。

Causal Models in Philosophy — Logic of counterfactuals

- ▶ 對於闡述虛擬情況的語句，我們心中有一套原則來判斷其真假。
 - ✓ 「假如我之前乖乖洗手，就不會拉肚子了」
 - ✗ 「假如我之前乖乖洗手，就會成為百萬富翁」
- ▶ 古典邏輯的“實質蘊涵” (material implication) 句，就是高中學的「若 p ，則 q 」。當我們知道語句 p 和 q 的真假時，就可以從真值表知道「若 p ，則 q 」整個句子的真假。
- ▶ 在實質蘊涵的真值表裡，當前件為假時，不管後件的真假如何，整個條件句都會為真。但這顯然與我們腦中存在的虛擬條件句語義理論並不相同，畢竟虛擬條件句的前件在現實世界多半都是假的，但我們卻不會因此認為後件隨便接什麼整句話都是真的。

Causal Models in Philosophy — Logic of counterfactuals

- ▶ Lewis 所提出的“最接近世界語義論” (closest-world semantics)：從“可能世界” (possible world) 的概念出發，發展出以比較世界之間的相似度來理解虛擬條件句的方法。
- ▶ Galles 與 Pearl 所提出的“結構模型語義論” (structural model semantics)：利用決定性因果模型、介入和子模型概念去理解虛擬條件句的理論。
- ▶ 在此簡單介紹這兩類語義理論的想法，我們使用“ $p > q$ ”來代表虛擬條件句「假如 p ，則 q 」的句型。

David Lewis: Closest-world Semantics

基於世界相似度的語義論是這樣理解虛擬條件句的：

1. 存在著一個關於 p 與 q 的現實世界，它是說出虛擬條件句的人所處的世界。
2. 除了現實世界之外，還存在許多可能世界，我們可以按照它們與現實世界的相似度來將它們加以排序，使其圍繞在現實世界周邊，越是類似於現實世界的可能世界距離現實世界就越接近。(在此必須先選定一個比較世界相似度的函數)
3. 虛擬條件句“ $p > q$ ”為真，若且唯若，最接近現實世界且 p 為真的世界們都是 q 為真的世界。

在第二步驟選擇比較可能世界相似度的函數時，排序是否為線性等細節會造成理論的差異。

David Lewis: Closest-world Semantics

使用前述的範例「假如我之前乖乖洗手，就不會拉肚子了」， p 為「我之前乖乖洗手」 q 為「不會拉肚子」，此時假設世界就由這兩個命題所構成：

1. w_a ：現實世界， p 為假， q 為假
2. w_1 ：超多病毒的可能世界， p 為真， q 為假，距離為 d_1
 w_2 ：有些病毒的可能世界， p 為真， q 為真，距離為 d_2
 w_3 ：沒有病毒的可能世界， p 為假， q 為真，距離為 d_3
3. 假設世界相似度定義在病毒含量與現實世界的差異，則距離的遠近關係可能是 $d_2 < d_1 < d_3$ 。因此， p 為真且距離最近的可能世界為 w_2 ，由於這也是一個 q 為真的世界，所以「假如我之前乖乖洗手，就不會拉肚子了」為真。

Galles and Pearl: Structural Model Semantics

基於結構模型的語義論是這樣理解虛擬條件句的：

1. 存在一個關乎 p 與 q 的模型 $\mathfrak{M}$ ，該模型反應了說出虛擬條件句的人所假設的因果機制。模型具有方向性，由外源性變數（包括 p ）和內源性變數（包括 q ），以及包括描述 p 與 q 之間關係的一組方程式所組成（假設描述 p 與 q 關係的方程式為 f ）。
2. 介入一個變數，會使原本方程式 f 的右側變成被介入的值，稱這個新的方程式為 f' ，稱新的模型為 $\mathfrak{M}'$ 。
3. 虛擬條件句“ $p > q$ ”為真，若且唯若， q 在所有介入前件 p 為真之後的每一個子模型中都為真。

Galles and Pearl: Structural Model Semantics

再次使用上面的例子， p 為「我之前乖乖洗手」 q 為「不會拉肚子」：

1. 存在一個模型，外源性變數 p 與內源性變數 q 的值域都是 $\{\text{真}, \text{假}\}$ ，相關的方程式為 $f: p = \text{假}; g: q = p$ 。
2. 介入 p 為真，則 $f': p = \text{真}$ 。
3. 在介入 p 為真的情況下得到 q 也為真的結果，因此「假如我之前乖乖洗手，就不會拉肚子了」為真。

這兩種語義理論間最大的差異，在於結構模型中的虛擬情況被認為是由“介入的實現”（realization by intervention）所造成，不管現實世界中的實際情況為何；也因為虛擬情況的真值條件完全被模型和介入所決定，因而並沒有相似度的概念在內。

威猛 Pearl 證明了 Lewis 的最接近世界語意論在公理系統上可以被他們的結構模型語意論推導出來。

Statistical vs Causal Models

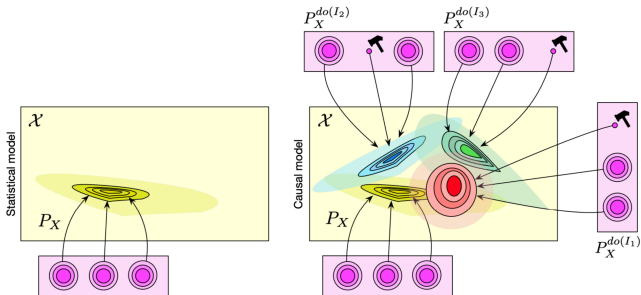


Fig. 1. Difference between statistical (left) and causal models (right) on a given set of three variables. While a statistical model specifies a single probability distribution, a causal model represents a set of distributions, one for each possible intervention (indicated with a $\blacktriangleright$ in the figure).

(Bernhard Schölkopf. Causality for Machine Learning. 2019)

Outline

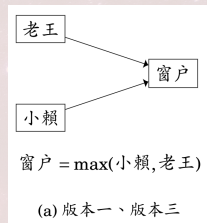
- ▶ A Brief History of Causal Models
 - ▶ Graphical model for association
 - ▶ Potential outcome framework
- ▶ **Causal Models in Philosophy**
 - ▶ Logic of counterfactuals
 - ▶ **Theories of actual causation**
- ▶ Causal Models in Machine Learning
 - ▶ Necessary and sufficient conditions
 - ▶ Causal discovery
 - ▶ Counterfactual learning for recommendation system
- ▶ Causal Models in Biostatistics and Epidemiology
 - ▶ Assumptions for causal inference
 - ▶ Mediation analysis

Theories of Actual Causation

根據前述 Hume 的看法，個例因果關係的成立與否似乎與某種虛擬條件句的真假息息相關，但它們之間的具體關係究竟如何呢？

► 丟石砸窗，版本一

1. 老王與小賴兩人都拿著石頭
2. 小賴朝著窗戶丟石頭，老王沒丟
3. 小賴的石頭砸破了窗戶



直覺上，小賴丟石頭（而非老王沒丟石頭）是窗戶破損的原因，但是，純粹從方程式或是因果圖上看來，小賴的位置與老王的位置是對稱的。那麼，是什麼事情使得這兩者之間具有因果關係呢？

Theories of Actual Causation — David Lewis

依照 Lewis 的理論，如果我們想要知道一件事 ($X = x$) 是否是另一件事 ($Y = y$) 的原因，我們就必須檢查它們是否滿足以下的條件：

AC1 在世界 w 裡， $X = x$ 為 $Y = y$ 的實際原因，若且唯若：

1. X 和 Y 為不同的變數
2. 在世界 w 裡， $X = x$ 且 $Y = y$
3. 存在某個 $x' \neq x$ 及某個 $y' \neq y$ ，使得 $X = x' > Y = y'$ 在 w 中為真

我們就可以發現“小賴丟石頭”是“窗戶破損”的原因，因為：

1. “小賴”和“窗戶”是不同的變數
2. 在現實世界裡，小賴丟了石頭且窗戶破損
3. 虛擬條件句“小賴沒丟石頭 $>$ 窗戶沒破”在現實世界中為真

我們可以發現“老王沒丟石頭”不是“窗戶破損”的原因，因為：

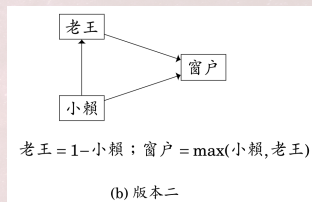
1. “老王”和“窗戶”是不同變數
2. 在現實世界裡，老王沒丟石頭且窗戶破損
3. 虛擬條件句“老王丟了石頭 $>$ 窗戶沒破”在現實世界中為假

Theories of Actual Causation — David Lewis

哲學家很快就發現了一些漏洞，常見的類型有“超前部署” (Preemption) 和“過度決定” (Overdetermination)。

► 丟石砸窗，版本二：“超前部署”

1. 老王和小賴兩人都拿著石頭
2. 老王決定先觀察小賴，如果小賴丟了石頭，他就不丟；但如果小賴不丟石頭，他就會丟石砸窗
3. 小賴朝著窗戶丟石頭
4. 小賴的石頭砸破了窗戶



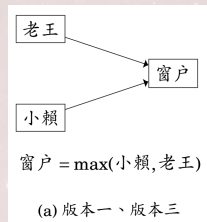
直覺上，小賴丟石頭仍然是窗戶破損的原因。但如果我們以 AC1 去評估這個故事，我們會發現“小賴丟了石頭”並不是“窗戶破損”的原因，因為：

1. “小賴”和“窗戶”是不同變數
2. 在現實世界裡，小賴丟了石頭且窗戶破損
3. 虛擬條件句“小賴沒丟石頭 > 窗戶沒破”在現實世界為假

Theories of Actual Causation — David Lewis

► 丟石砸窗，版本三：“過度決定”

1. 老王與小賴兩人都拿著石頭
2. 兩人同時朝著窗戶丟石頭
3. 兩人的石頭同時砸破了窗戶



直覺上，任意一人丟石頭都是為窗戶破損的原因。但如果我們以 AC1 去評估這個故事，我們會發現“小賴丟了石頭”和“老王丟了石頭”都不是“窗戶破損”的原因，因為：

1. “小賴”和“窗戶”是不同變數
2. 在現實世界裡，小賴丟了石頭且窗戶破損
3. 虛擬條件句“小賴沒丟石頭 > 窗戶沒破”在現實世界為假

在版本二跟版本三的故事中，我們的直覺都是：“小賴丟了石頭”是窗戶破損的現實原因，但依照 Lewis 的理論卻不是。

Theories of Actual Causation — Joseph Halpern

有許多的修正理論被提出，在此簡單介紹 Halpern 的理論：

AC2 在世界 w 裡， $X = x$ 為 $Y = y$ 的實際原因，若且唯若：

1. X 和 Y 是不同變數
2. 在世界 w 裡， $X = x$ 且 $Y = y$
3. 存在兩個不相交的集合 X 和 Z ，且 $X \in \mathbf{X}$ 。在世界 w 中， $\mathbf{X} = \mathbf{x}$ 且 $\mathbf{Z} = \mathbf{z}$ ，使得：
 - 3.1 存在 $\mathbf{x}' \neq \mathbf{x}$ 使得 “ $(X = \mathbf{x}' \& Z = \mathbf{z}) > Y \neq y$ ” 在世界 w 中為真
 - 3.2 $\mathbf{X}$ 中沒有任何子集滿足條件 3.1

AC1 第三個條件中提到的虛擬條件句的前件只關心 X 值的改變，但 AC2 的第三個條件則在虛擬條件句的前件還額外固定了‘背景變數’ Z 的實際值。

Theories of Actual Causation — Joseph Halpern

以超前部署為例，AC2 判定“小賴丟了石頭”是窗戶破損的原因，因為：

1. “小賴”和“窗戶”是不同變數
2. 在現實世界裡，小賴丟了石頭且窗戶破損
3. 當我們以“老王沒丟石頭”作為被控制的背景變數，則虛擬條件句“小賴沒丟石頭且老王也沒丟石頭 $\succ$ 窗戶沒破”在現實世界中為真。

再以過度決定為例，AC2 判定“小賴丟了石頭”和“老王丟了石頭”都是“窗戶破損”的原因，因為：

1. “小賴”和“窗戶”是不同變數
2. 在現實世界裡，小賴丟了石頭且窗戶破損
3. 如果我們讓 $X = \{\text{小賴}, \text{老王}\}$ 而讓 $Z = \emptyset$ ，那麼，我們可以很容易看出虛擬條件句“小賴沒丟石頭且老王沒丟石頭 $\succ$ 窗戶沒破”在現實世界中為真，但 $\{\text{小賴}\}$ 和 $\{\text{老王}\}$ 則都不符合 AC2 中 3(a) 的條件。

Theories of Actual Causation

- ▶ 從 AC1 到 AC2 可以看出：在討論個例因果時，除了“因”與“果”之外，我們可能還得關心“背景”的因素。
- ▶ 類似的情況也出現在規律性分析的理論中。早期提倡規律性分析理論的學者使用充分條件與必要條件來定義個例因果關係，但最後發展出的理論往往需要定義一個“因果場”來設定背景。如此一來，因果關係的認定逐漸變成相對性的，是針對某個模型所下的結論。

Outline

- ▶ A Brief History of Causal Models
 - ▶ Graphical model for association
 - ▶ Potential outcome framework
- ▶ Causal Models in Philosophy
 - ▶ Logic of counterfactuals
 - ▶ Theories of actual causation
- ▶ **Causal Models in Machine Learning**
 - ▶ **Necessary and sufficient conditions**
 - ▶ Causal discovery
 - ▶ Counterfactual learning for recommendation system
- ▶ Causal Models in Biostatistics and Epidemiology
 - ▶ Assumptions for causal inference
 - ▶ Mediation analysis

Sufficient condition: Markov condition / MC

- ▶ 圖形學派希望相關性就可以代表因果關係，所以充分條件為“有相關就有因果”或“沒因果就沒相關”，必要條件為“有因果就有相關”。
- ▶ 假設外源性變數之間是獨立的，Pearl 與 Verma 證明內源性變數的分佈就會滿足 Markov 條件，這裡介紹幾種常見的寫法：
 - ▶ $MC_{\text{Screening_off}} : P(X|PA(X) \& Y) = P(X|PA(X))$
 - ▶ $MC_{\text{Factorization}} : P(X_1, X_2, \dots, X_n) = \prod P(X_i|PA(X_i))$
 - ▶ $MC_{\text{d-separation}} : d$ 是 directional，d-separation 其實就是“此路不通”的意思。若在圖 G 中 Z 使 X 與 Y 不相關，則 $P(X, Y|Z) = P(X|Z) \times P(Y|Z)$ 。

Sufficient condition: Markov condition / MC

SEP 上條列了三種 MC 可能不成立的情況：

- ▶ 量子糾纏：假設 X 為一粒子自旋方向， Y 為另一粒子的自旋方向，且兩粒子呈糾纏態。當 X 被測量時，我們即時就知道 Y 的測量值，且在實驗中可以保證 X 跟 Y 沒有共同原因。所以就資料上來看， X 與 Y 會出現非常強的關聯性，但卻沒有因果關係存在（前提是我們預設因果關係是一種物理世界可以傳送的訊號，詳見 EPR 實驗說明）。
- ▶ 存在沒觀察到的干擾變數：假設 X 、 Y 與 Z 為統計獨立之變數，且無因果關係。今天如果因果模型內沒有這三個變數，卻有 $U = X + Y$ 及 $W = Y + Z$ ，則變數 U 和 W 會因為沒有被觀察到的干擾變數 Y 的存在而出現相關性，但此相關性並非來自因果關係。
- ▶ 資料不精確：假設 X 、 Y 與 Z 為連續的變數， Z 為 X 與 Y 之共同原因且 X 與 Y 之間並無因果關係。今天如果因果模型中沒有 Z ，只有該變數的簡略版 Z' ，只指出了原變數的數值是高還是低。如此我們就無法只用 Z' 來讓 X 與 Y 獨立，因為 Z 的資訊不完整，即使在給定 Z' 的情況下， X 跟 Y 在數據上看起來依然會相關。

這三個情況都是“沒因果但有相關”。

Necessary condition: Faithfulness condition / FC

- ▶ MC 要求 d-separated 時路線兩端的變數獨立，FC 才保證了 d-connected 的變數絕對在統計上相關：
 - ▶ 極簡條件 (Minimality Condition)：假設因果圖 G 裡變數的集合 V 在機率分佈 P 之下滿足了 MC，極簡條件要求沒有 G 的子圖，是擁有同樣的變數 V 與機率 P 但卻又滿足 MC 的。換句話說， G 裡所有的邊都是必要的，且有邊連著就代表兩變數統計上相關。
 - ▶ 忠實條件 (Faithfulness Condition, FC)：忠實條件比極簡條件更嚴格，忠實條件要求變數之間若是獨立，都必須是 d-separated。換句話說，極簡條件要求有邊連著就有相關性，忠實條件要求有路徑連著就得有相關性，也就保證了所有 d-connected 的變數都在統計上相關。
- ▶ 忠實條件在“完美相消” (perfect cancellation) 的情況之下不成立。也就是說變數間有多條路線，但這些路線加起來卻剛好互相抵消，使得計算時變數之間的相關性剛好消失，也就是“沒相關但有因果”的情況。

Necessary and sufficient conditions

- ▶ 圖形模型的基本化約單位就是相關性，上述所有的條件都是希望相關性就能代表因果關係，所以我們極盡可能的把“沒因果但有相關”及“有因果但沒相關”都排除掉，前者使用 MC，後者使用 FC。
- ▶ 假設了這些條件之後，我們就可以說因果圖上有相關性的變數都具有因果性，但這是在因果圖已知的情況下，而要決定因果圖的內容，實務上大都還是依賴相關領域的知識。如果想要完全從資料來“猜出”因果結構，我們又需要一些其他的假設及演算法。

Outline

- ▶ A Brief History of Causal Models
 - ▶ Graphical model for association
 - ▶ Potential outcome framework
- ▶ Causal Models in Philosophy
 - ▶ Logic of counterfactuals
 - ▶ Theories of actual causation
- ▶ **Causal Models in Machine Learning**
 - ▶ Necessary and sufficient conditions
 - ▶ **Causal discovery**
 - ▶ Counterfactual learning for recommendation system
- ▶ Causal Models in Biostatistics and Epidemiology
 - ▶ Assumptions for causal inference
 - ▶ Mediation analysis

Causal discovery — Judea Pearl

- ▶ 在已知因果圖的情況下，假設 MC 跟 FC 成立就能讓因果關係化約成“帶有方向的相關性”。但如果能夠由資料來反推因果關係，那對於因果推論的自動化將是一大助益。
- ▶ 因果發現的問題定義如下：令資料內變數的集合為 V ，且已知變數的機率分佈為 P ，如何決定哪些變數間有邊相連及其方向？有沒有邊相連，可以用條件機率來檢查相關性，而邊的方向則需要另外的資訊。
- ▶ 例如 Pearl 在 1988 年提出辨識因果的條件為“時間順序”的資訊：
 - ▶ V 裡的變數以時間排序，時間早的才有可能成為時間晚的之原因。
 - ▶ 不論 V 裡的變數被設定成任何可能的值，對應的機率分佈 P 都能夠給予該變數一個合法的機率。
 - ▶ 模型內不存在“潛在變數”，所以對應的因果圖 G 是一個 DAG。
 - ▶ G 裡變數們的機率分佈 P ，得滿足 MC 以及極簡條件。

如果資料滿足上述條件，則可以從 P 來反推出唯一的因果圖 G 。

Causal discovery — Judea Pearl

推論的演算法如下：

1. 首先依照時間順序把變數排序，時間早的為時間晚的父點候選人。

$$\underbrace{X_1, \dots, X_{i-1}}_{\text{父點候選人}}, X_i, \underbrace{X_{i+1}, \dots, X_n}_{\text{子點候選人}}$$

父點候選人如果無法成為父點，因為時間順序的關係也不可能成為子點，從 MC 我們知道，這樣的變數會被父點屏蔽。

- 1.1 針對變數 X_i ，先計算條件機率 $P(X_i | X_1, \dots, X_{i-1})$ ，然後嘗試把父點候選人移除，看看條件機率是否會有所改變。那些移除後不影響條件機率的，依據極簡條件我們知道不可能是 X_i 的父點。
- 1.2 那些的確會影響條件機率的變數，即為 X_i 的父點。
2. 使用上述的規則，從 X_1 一路計算到 X_n ，就可以畫出一張唯一的因果圖 G ，因果圖的邊來自於相關性，而方向來自於時間順序。

Causal discovery — Markov equivalence class

- ▶ 如果沒有時間順序或是其他能夠排序變數的規則，單從條件機率加上 MC 跟 FC 無法完全決定因果的方向性，在這種情況下做為答案的因果圖也就不是唯一解，而是符合同樣條件機率分佈的圖之集合，稱之為因果結構上的“Markov 等價分類” (Markov equivalence class, MEC)，Pearl 稱這種因果結構上等價的圖形表示法為“模式” (pattern)。
- ▶ 舉例，假如模型中有三個變數 $\{X, Y, Z\}$ ，且只有在給定 Y 時 X 跟 Z 會獨立，其他情況下變數都相關時，我們發現符合這樣條件的因果圖有：

$$X \rightarrow Y \rightarrow Z, \quad X \leftarrow Y \leftarrow Z, \quad X \leftarrow Y \rightarrow Z$$

條件機率加上 MC 跟 FC 也無法決定這三張因果圖哪一個是真的，只排除了 Y 作為對撞變數的情況，因此這三張因果圖就被視為一個 MEC。

Causal discovery — Recent development

因果發現演算法近來蓬勃發展，依照 Glymour 在 2019 年寫的一篇回顧，這些方法可以分成兩大類：

- ▶ 經典的“限制及計分方法”，包含了有名的 PC 跟 FCI 演算法以及應用貪婪搜尋來找尋 MEC 的作法；
- ▶ 使用更多統計假設的“函數因果模型”，此類方法額外加入了機率分佈假設以求獲得比 MEC 更加精確的因果圖。

要注意兩點：其一，這套方法的因果關係不是建立在比較虛擬情況上，而是化約成接受 MC 跟 FC 的相關性；其二，因為這套方法是從資料來找出符合 MC 跟 FC 的因果模式，比較適合變數多資料量也多的情形，不然得到的結果可能很不準確。

Outline

- ▶ A Brief History of Causal Models
 - ▶ Graphical model for association
 - ▶ Potential outcome framework
- ▶ Causal Models in Philosophy
 - ▶ Logic of counterfactuals
 - ▶ Theories of actual causation
- ▶ **Causal Models in Machine Learning**
 - ▶ Necessary and sufficient conditions
 - ▶ Causal discovery
 - ▶ **Counterfactual learning for recommendation system**
- ▶ Causal Models in Biostatistics and Epidemiology
 - ▶ Assumptions for causal inference
 - ▶ Mediation analysis

Counterfactual learning for recommendation system

God's viewpoint 

id	vaccine	COVID
A	true	Y
B	true	Y
C	false	Y
D	false	N
E	false	N

actual world

id	vaccine	COVID
A	true	Y
B	true	Y
C	true	Y
D	true	Y
E	true	N

vaccinated world

id	vaccine	COVID
A	false	Y
B	false	N
C	false	Y
D	false	N
E	false	N

naïve world

Counterfactual learning for recommendation system

The image shows a Bing search results page for the query "organic apples". The search bar at the top shows "organic apples" and "bing" logo. Below the search bar, it says "100,000,000 RESULTS". The main content area displays several search results. Two specific areas are highlighted with red dashed boxes and callouts:

- Mainline:** A red dashed box highlights the first search result: [Organic Just Apples](#) from [iHerb.com](#). The snippet reads: "Consumer Rated #1 Online Retailer - Great Value and Fast Shipping (iHerb.com is rated on PriceGrabber (43 reviews))".
- Sidebar:** A red dashed box highlights an advertisement in the right sidebar. The ad is titled "Organic Fruit Deal \$29.99" and includes the text: "www.CherryMoonFarms.com/Fruit Use PromoCode GET10 for Discount on All Fresh Organic Fruit Baskets cherrymoonfarms.com is rated on Bizrate (106 reviews)". Below this, another ad for "Organic Fruit Delivery" from TheFruitCompany.com is visible, followed by "Organic Apples at Amazon" with the text "Low prices on Organic Apples. Qualified orders over \$25 ship free".

Figure 1: Mainline and sidebar ads on a search result page. Ads placed in the mainline are more likely to be noticed, increasing both the chances of a click if the ad is relevant and the risk of annoying the user if the ad is not relevant.

source: Counterfactual Reasoning and Learning Systems: The Example of Computational Advertising

Counterfactual learning for recommendation system

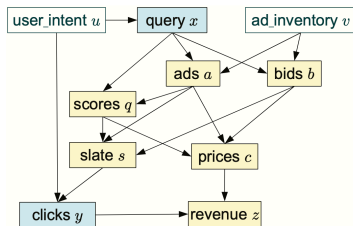


Figure 3: Causal graph associated with the structural equation model of Figure 2. The mutually independent noise variables ϵ_1 to ϵ_8 are implicit. The variables a , b , q , s , c , and z depend on their direct causes in known ways. In contrast, the variables u and v are exogenous and the variables x and y depend on their direct causes through unknown functions.

Counterfactual learning for recommendation system

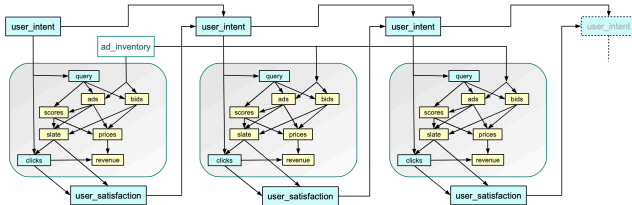


Figure 4: Conceptually unrolling the user feedback loop by threading instances of the single page causal graph (Figure 3). Both the ad slate s_t and user clicks y_t have an indirect effect on the user intent u_{t+1} associated with the next query.

Outline

- ▶ A Brief History of Causal Models
 - ▶ Graphical model for association
 - ▶ Potential outcome framework
- ▶ Causal Models in Philosophy
 - ▶ Logic of counterfactuals
 - ▶ Theories of actual causation
- ▶ Causal Models in Machine Learning
 - ▶ Necessary and sufficient conditions
 - ▶ Causal discovery
 - ▶ Counterfactual learning for recommendation system
- ▶ **Causal Models in Biostatistics and Epidemiology**
 - ▶ **Assumptions for causal inference**
 - ▶ Mediation analysis

Assumptions for causal inference: Exchangeability

- ▶ 可交換性 (exchangeability)：使用對方群體在現實世界的機率來猜測自己群體在虛擬情況的機率。

$$P(\text{肺癌}^{\text{抽菸}} | \text{不抽菸}) = P(\text{肺癌}^{\text{抽菸}} | \text{抽菸})$$

「強迫真實世界不抽菸的人在虛擬情況抽菸，這群人得到肺癌的機率，與真實世界抽菸的人得到肺癌的機率是一樣的」

$$P(\text{肺癌}^{\text{不抽菸}} | \text{抽菸}) = P(\text{肺癌}^{\text{不抽菸}} | \text{不抽菸})$$

「強迫真實世界抽菸的人在虛擬情況不抽菸，這群人得到肺癌的機率，與真實世界不抽菸的人得到肺癌的機率，是一樣的」

Assumptions for causal inference: Consistency

- ▶ 不管你在真實世界抽煙與否，如果你掉進了一個強迫抽菸的虛擬情況，你得到肺癌的機率就是固定是 $P(\text{肺癌}^{\text{抽菸}})$ ；而掉進不抽菸的虛擬情況，得到肺癌的機率固定是 $P(\text{肺癌}^{\text{不抽菸}})$ ，因此：

$$P(\text{肺癌}^{\text{抽菸}}) = P(\text{肺癌}^{\text{抽菸}} | \text{不抽菸}) = P(\text{肺癌}^{\text{抽菸}} | \text{抽菸})$$

$$P(\text{肺癌}^{\text{不抽菸}}) = P(\text{肺癌}^{\text{不抽菸}} | \text{抽菸}) = P(\text{肺癌}^{\text{不抽菸}} | \text{不抽菸})$$

- ▶ $P(\text{肺癌}^{\text{抽菸}} | \text{不抽菸})$ 就是我們缺失的、無法觀察的虛擬情況。
- ▶ $P(\text{肺癌}^{\text{抽菸}} | \text{抽菸})$ 雖然也使用虛擬情況的定義，但看起來與條件機率非常相似：「強迫真實世界抽菸的人在虛擬情況抽菸得到肺癌的機率」
- ▶ 因為這個虛擬情況跟現實世界所描述的行為相同，所以我們直覺認為這句話可以讀成「真實世界抽菸的人得到肺癌的機率」，這個想法被稱為“一致性” (consistency)，或著 Pearl 稱之為“構成” (composition)。

Assumptions for causal inference

- ▶ 一致性是連結虛擬情況與現實世界必要的假設，也是讓虛擬機率能夠被計算的依據；但它也是一個無法被驗證的假設，畢竟身在現實世界的我們無法觀察虛擬情況。
- ▶ 注意上述我並沒有說該虛擬情況跟現實世界相等，畢竟隨機實驗的根本就是給定一樣的條件也有可能出現不同結果。因此，在無法確認虛擬情況跟現實世界等同的前提下，一致性假設他們至少在機率上相等：

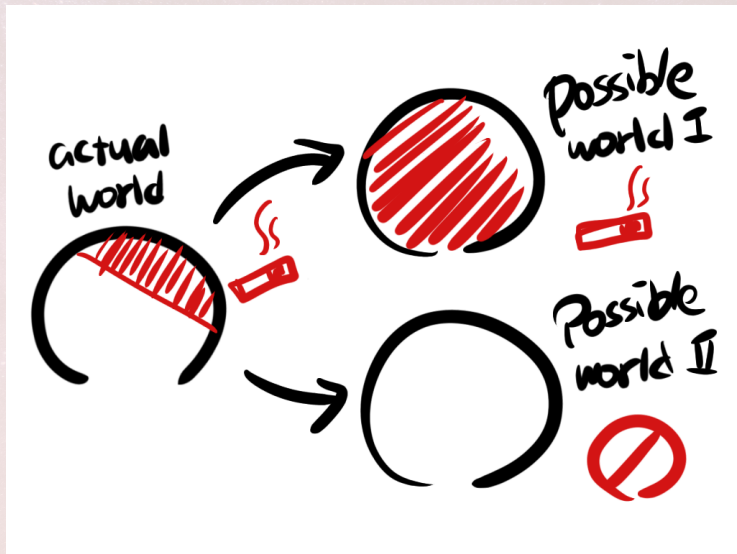
$$P(\text{肺癌}^{\text{抽菸}} | \text{抽菸}) = P(\text{肺癌} | \text{抽菸})$$

$$P(\text{肺癌}^{\text{不抽菸}} | \text{不抽菸}) = P(\text{肺癌} | \text{不抽菸})$$

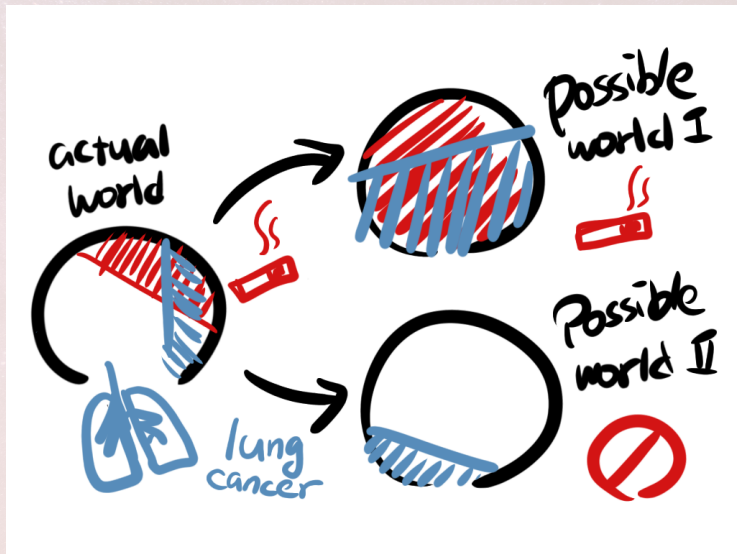
- ▶ 可以看出有了一致性之後，虛擬機率就可以使用條件機率來計算了：

$$\begin{aligned} P(\text{肺癌}^{\text{抽菸}}) &= P(\text{肺癌}^{\text{抽菸}} | \text{抽菸}) \quad \boxed{\text{可交換性}} \\ &= P(\text{肺癌} | \text{抽菸}) \quad \boxed{\text{一致性}} \end{aligned}$$

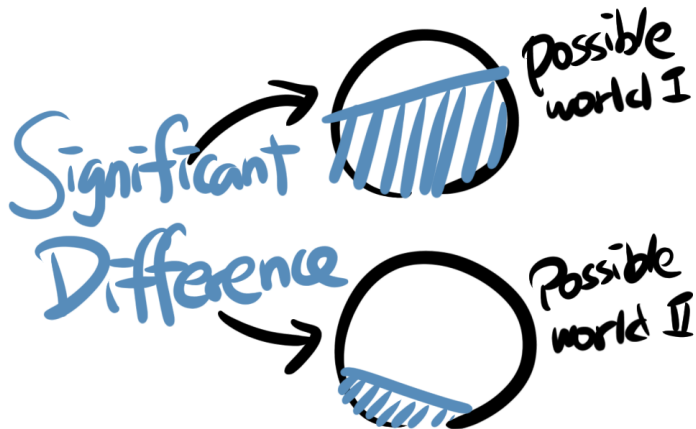
Assumptions for causal inference



Assumptions for causal inference



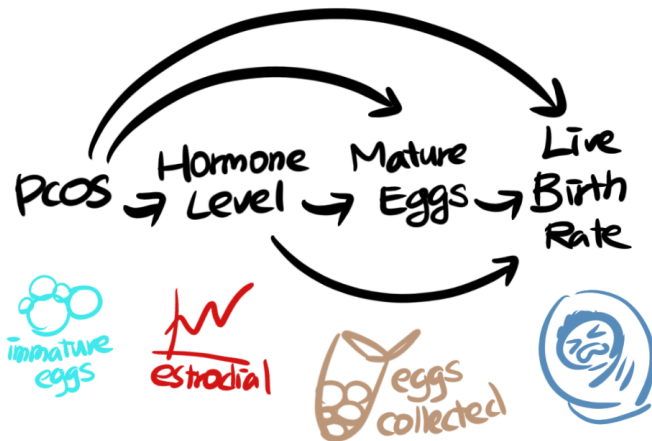
Assumptions for causal inference



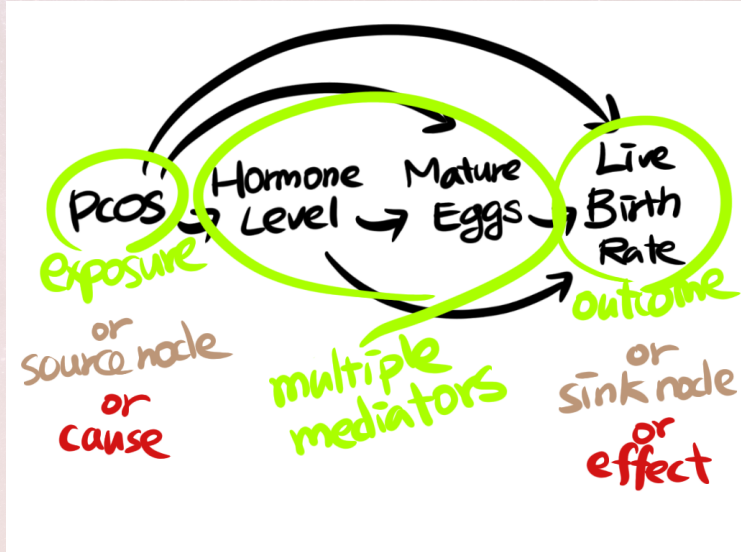
Outline

- ▶ A Brief History of Causal Models
 - ▶ Graphical model for association
 - ▶ Potential outcome framework
- ▶ Causal Models in Philosophy
 - ▶ Logic of counterfactuals
 - ▶ Theories of actual causation
- ▶ Causal Models in Machine Learning
 - ▶ Necessary and sufficient conditions
 - ▶ Causal discovery
 - ▶ Counterfactual learning for recommendation system
- ▶ **Causal Models in Biostatistics and Epidemiology**
 - ▶ Assumptions for causal inference
 - ▶ **Mediation analysis**

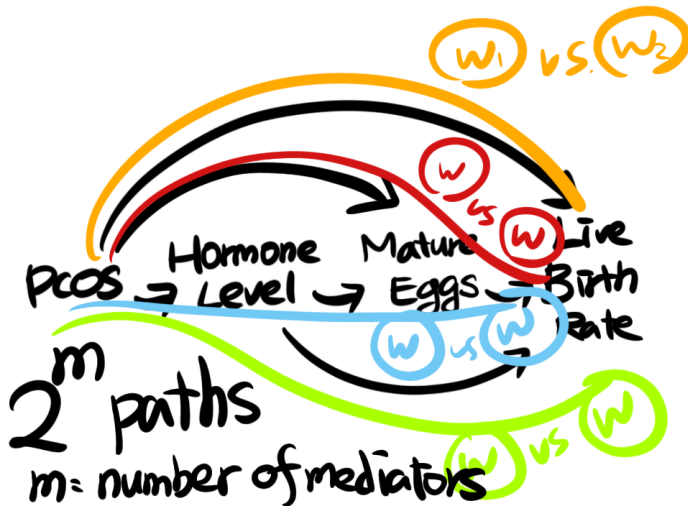
Mediation Analysis



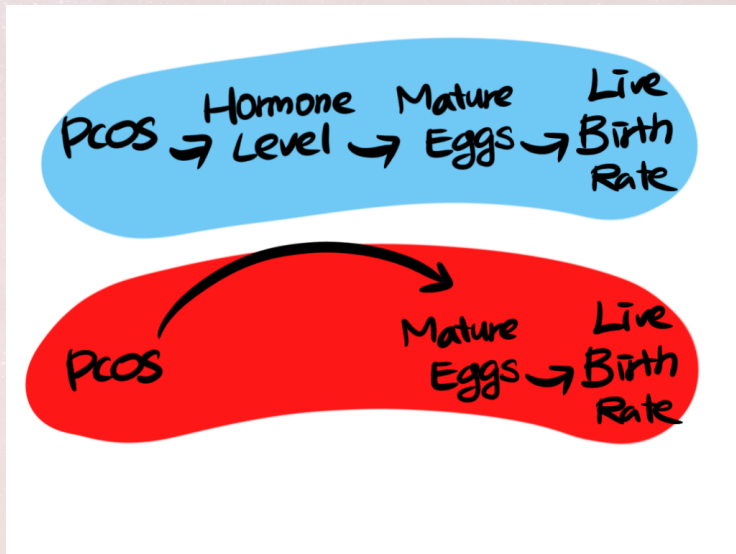
Mediation Analysis



Mediation Analysis



Mediation Analysis



Advertisement



← In case you are
interested....

Thank You

