An Interventionist Approach to Mediation Analysis

报告人：胡文杰
北京大学数学科学学院博士生
huwenjie@pku.edu.cn

基础信息

- 该论文的基础信息
  - 论文作者：James Robins
  - 提出了FFRCISTG，G-formula等工作，在因果推断领域有很大贡献。

James M. Robins

FOLLOW

professor of epidemiology and biostatistics harvard school of public health
Verified email at hsph.harvard.edu

TITLE	CITED BY	YEAR
Marginal structural models and causal inference in epidemiology JM Robins, MA Hernan, B Brumback Epidemiology 11 (5), 550-560	3824	2000
Causal diagrams for epidemiologic research S Greenland, J Pearl, JM Robins Epidemiology, 37-48	2714	1999
Estimation of regression coefficients when some regressors are not always observed JM Robins, A Rotnitzky, LP Zhao Journal of the American statistical Association 89 (427), 846-866	2210	1994
Causal diagrams for empirical research J Pearl Biometrika 82 (4), 669-688	1848	1995
A structural approach to selection bias MA Hernán, S Hernández-Díaz, JM Robins Epidemiology, 615-625	1813	2004
A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect J Robins Mathematical modelling 7 (9-12), 1393-1512	1748	1986
Analysis of semiparametric regression models for repeated outcomes in the presence of missing data JM Robins, A Rotnitzky, LP Zhao Journal of the american statistical association 90 (429), 106-121	1525	1995

论文背景

- Robins认为Pearl的NPSEM-IE(nonparametric structure equation model with independent errors)的模型对于counterfactual因果模型的描述不够精确。该文章主要详细阐述了中介分析的一种新视角。

- Robins认为自己在该论文中提出的interventionist方法在实际中更能被各个领域的专家所接受。同时也能更好的看出中介分析识别性的一些假设。

以往的进展

- 该文章主要基于Robins的FFRCISTG(finest fully randomized causal interpreted structural tree model)/SWIG(Single World Intervention Graph)以及Pearl的NPSEM-IE。关于中介分析，之前往往关心下面一些参数。

- 如果将处理A和中介变量都看成处理，则可定义CDE(controlled direct effect):

- $CDE_{a,a'}(m) = E[Y(a',m) - Y(a,m)]$

- 另一个则是PDE(pure direct effect)或NDE(natural direct effect)

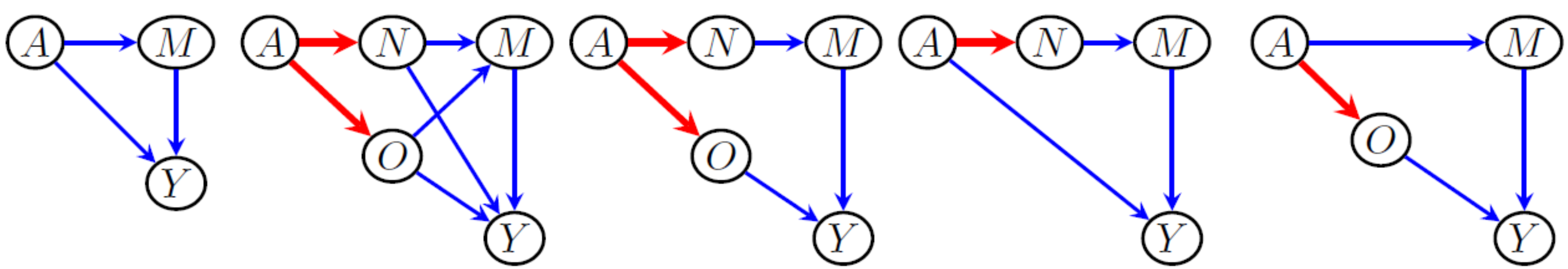
- $PDE_{a,a'} = E[Y\left(a',M(a)\right) - Y\left(a,M(a)\right)]$

解决的科学问题

- 先分析了CDE和PDE的识别性问题。
- 然后提出了基于干预的中介分析新框架。
- 并给出了识别性的一些定理。

解决问题的思路

- Robins在该文章中的想法是：将处理A分解成两部分N和O，其中N对中介变量M有作用，而O对结局变量Y有作用。



(a) (b) (c) (d) (e)

背景知识

- Pearl的图模型与Robins 的FFRCISTG模型

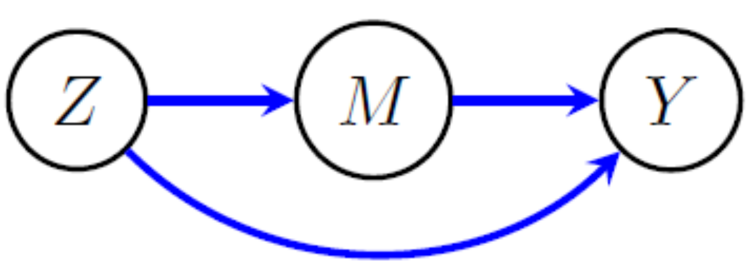
Definition 1 (NPSEM Counterfactual Existence Assumption)**.**

(i) *For each variable* $V \in \mathbf{V}$ *and assignment* $\widetilde{\mathbf{pa}}$ *to* $\mathrm{pa}_{\mathcal{G}}(V)$*, the parents of* V *in* $\mathcal{G}$*, we assume the existence of a counterfactual variable* $V(\widetilde{\mathbf{pa}})$*.*

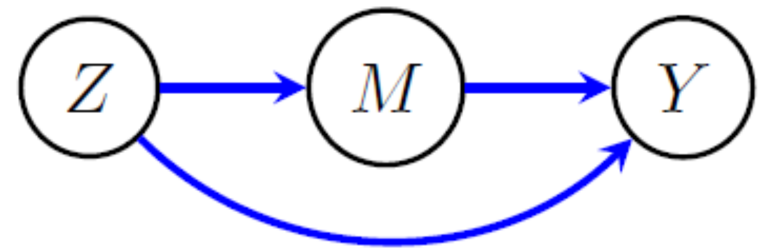
(ii) *For any set* $\mathbf{R}$*, with* $\mathbf{R} \neq \mathrm{pa}_{\mathcal{G}}(V)$*,* $V(\tilde{\mathbf{r}})$ *is defined recursively via:*

$$V(\tilde{\mathbf{r}}) = V\left(\tilde{\mathbf{r}}_{(\mathrm{pa}_{\mathcal{G}}(V)\cap\mathbf{R})}, (\mathbf{PA}_V \setminus \mathbf{R})(\tilde{\mathbf{r}})\right), \tag{14}$$

where $(\mathbf{PA}_V \setminus \mathbf{R})(\tilde{\mathbf{r}}) \equiv \{V^*(\tilde{\mathbf{r}}) \mid V^* \in \mathrm{pa}_{\mathcal{G}}(V), V^* \notin \mathbf{R}\}$.



背景知识



- Pearl的图模型与Robins 的FFRCISTG模型

Definition 2 (FFRCISTG Independence Assumption[27]).

For every $\mathbf{v}^\dagger$, *the variables* $\left\{ V(\mathbf{pa}_V^\dagger) \;\middle|\; V \in \mathbf{V},\ \mathbf{pa}_V^\dagger = \mathbf{v}^\dagger_{\mathrm{pa}_{\mathcal{G}}(V)} \right\}$ (17)

are mutually independent.

The NPSEM-IE model is the submodel of the FFRCISTG obeying the stronger independence assumption:[28]

Definition 3 (NPSEM-IE Independence Assumption).
The variables $\{\epsilon_V \mid V \in \mathbf{V}\}$ *are mutually independent.*
This is equivalent to:

The sets of variables $\left\{ \{V(\mathbf{pa}_V^\dagger) \mid \textit{for all } \mathbf{pa}_V^\dagger\} \;\middle|\; V \in \mathbf{V} \right\}$ (18)

are mutually independent.

背景知识

- 例子

- 在FFRCISTG模型下，我们假设

$$Z \perp M(z) \perp Y(z,m)$$

- 而在NPSEM-IE模型下，我们假设

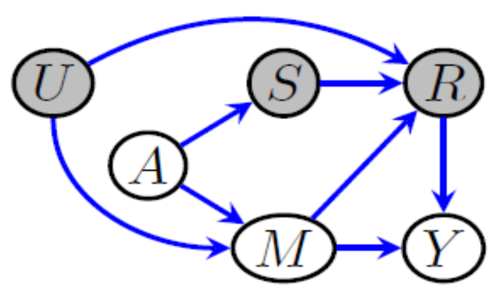
$$Z \perp \{M(z=0), M(z=1)\} \perp \{Y(z,m)\ for\ all\ z,m\}$$

- 例如在NPSEM-IE模型假设中，有$M(1) \perp Y(0,m)$，而该条件在FFRCISTG模型中并不一定成立。

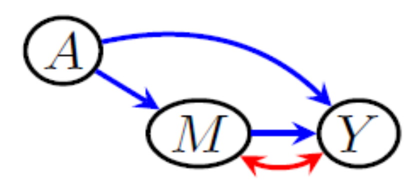
- 例子
- 在FFRCISTG模型下，我们假设 $Z \perp M(z) \perp Y(z,m)$
- 而在NPSEM-IE模型下，我们假设 $Z \perp \{M(z=0), M(z=1)\} \perp \{Y(z,m)\ for\ all\ z,m\}$
- 例如在NPSEM-IE模型假设中，有$M(1) \perp Y(0,m)$，而该条件在FFRCISTG模型中并不一定成立。

背景知识

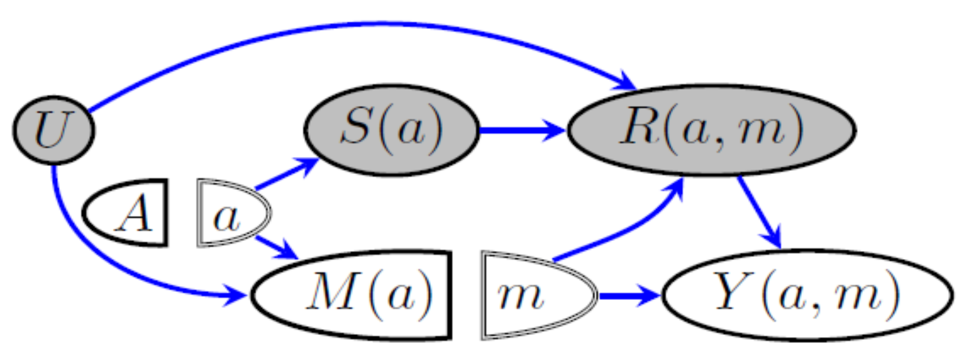
- Robins and Richardson的SWIG图模型，是一种表示方式，能够更直观的看出变量之间的相关关系。



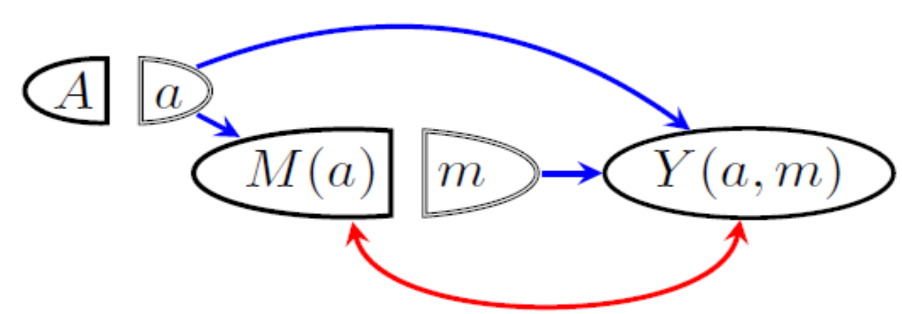
(a)



(a*)



(b)



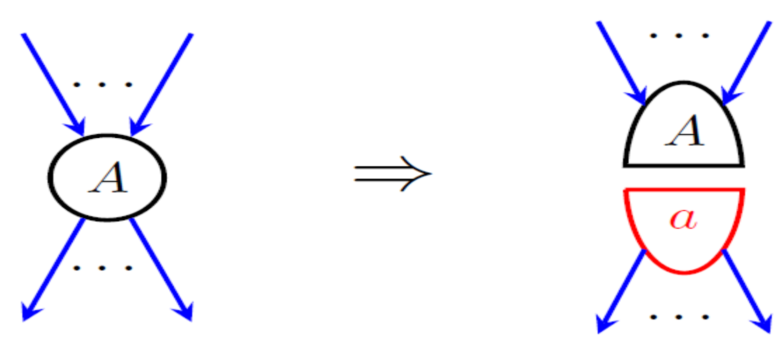
(b^*)

背景知识

- 构造SWIG的方法。

The SWIT $\mathcal{G}(\mathbf{a})$ resulting from intervening to set the variables in $\mathbf{A}$ to $\mathbf{a}$ in a directed acyclic graph $\mathcal{G}$ with vertex set $\mathbf{V}$ is constructed in two steps as follows:

(1) *Split Nodes*: For every $A \in \mathbf{A}$ split the node into a random and fixed component, labelled A and a respectively, as follows:



Splitting: Schematic Illustrating the Splitting of Node A

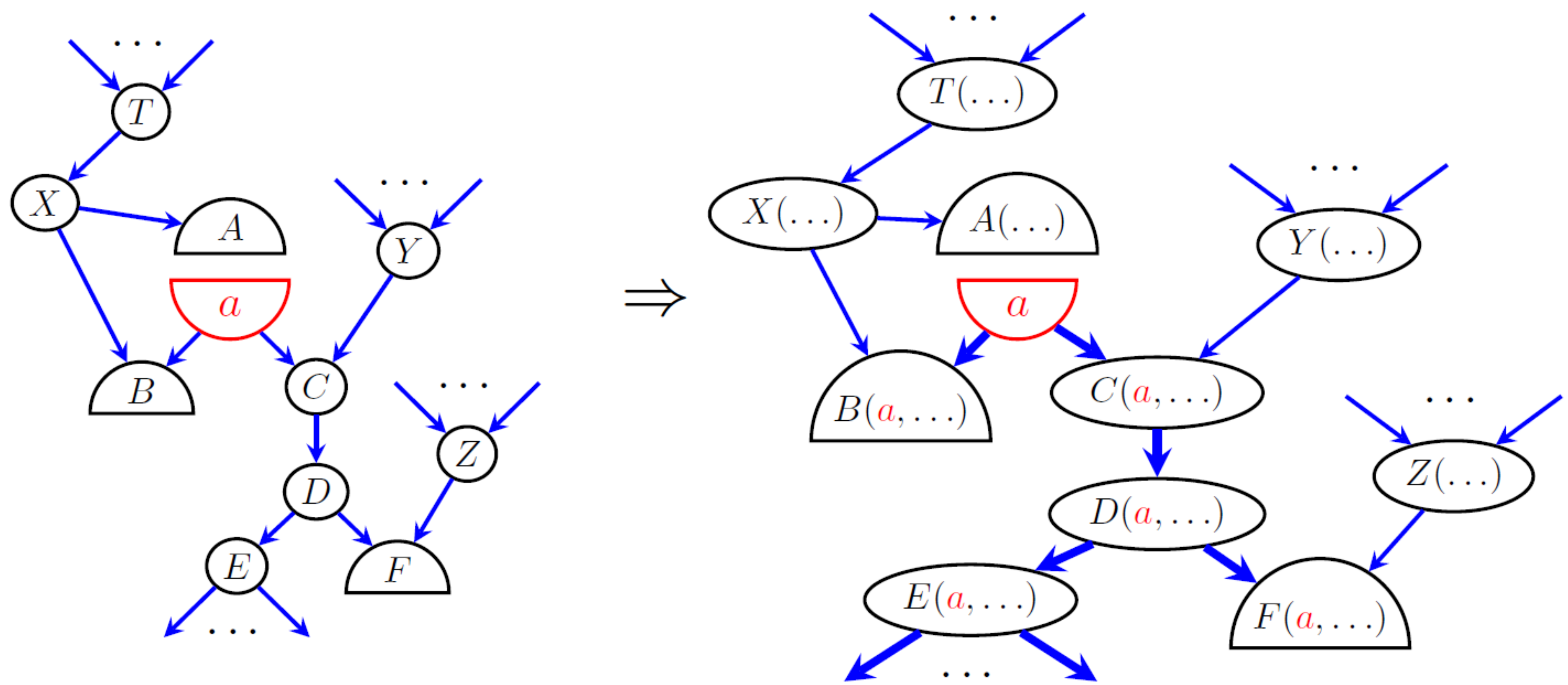
Thus the random half inherits all edges directed into A in $\mathcal{G}$; the fixed half inherits all edges directed out of A.

Let the resulting graph be $\mathcal{G}^*$. For each random vertex V in $\mathcal{G}^*$, let $\mathbf{a}_V$ denote the subset of fixed vertices that are ancestors of V in $\mathcal{G}^*$.

背景知识

(2) *Labeling:* For every random node V in $\mathcal{G}^*$, label it with $V(\mathbf{a}_V)$ (see the schematic below).

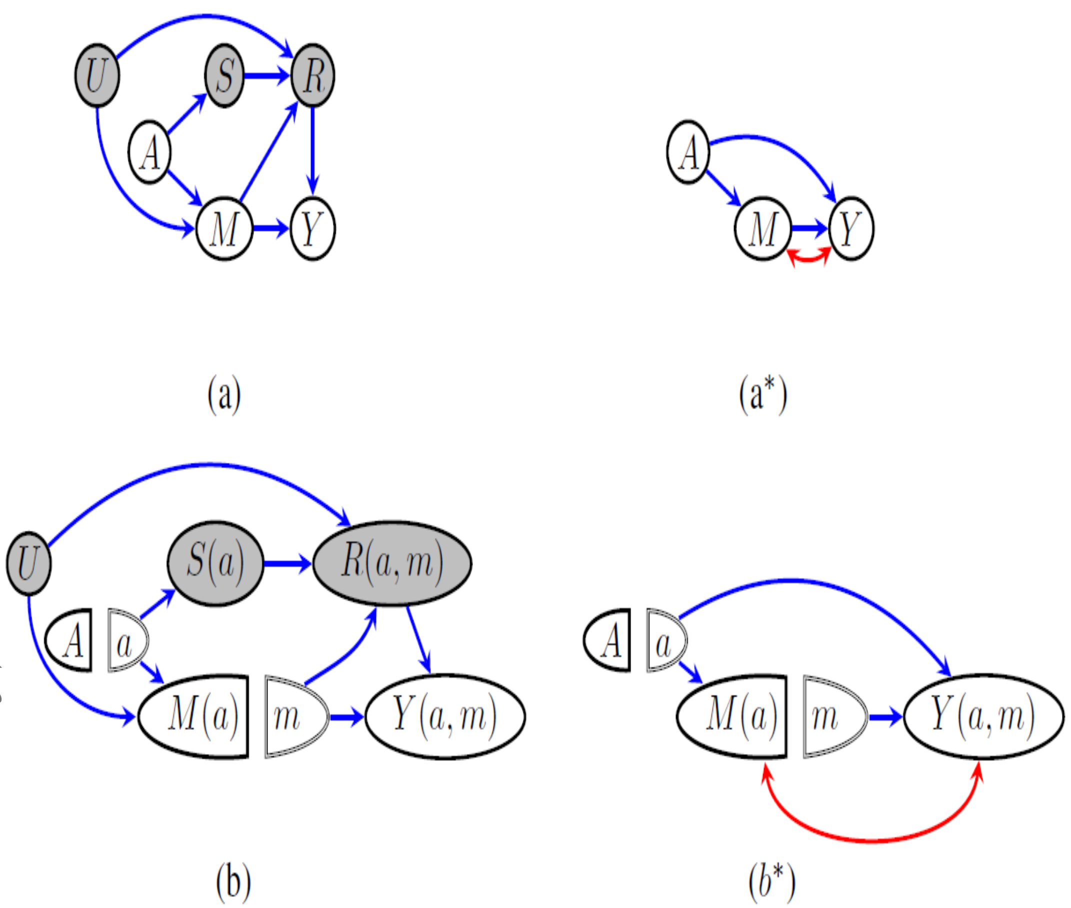
It is implicit here that if $\mathbf{a}_V = \emptyset$ then $V(\mathbf{a}_V) = V$. The resulting graph is the SWIT $\mathcal{G}(\mathbf{a})$. Let $\mathbb{V}(\mathbf{a}) \equiv \{V(\mathbf{a}_V) \mid V \in \mathbf{V}\}$ be the set of random vertices in $\mathcal{G}(\mathbf{a})$.



labeling: Schematic showing the nodes $V(\mathbf{a}_V)$ in $\mathcal{G}(\mathbf{a})$ for which $a \in \mathbf{a}_V$.

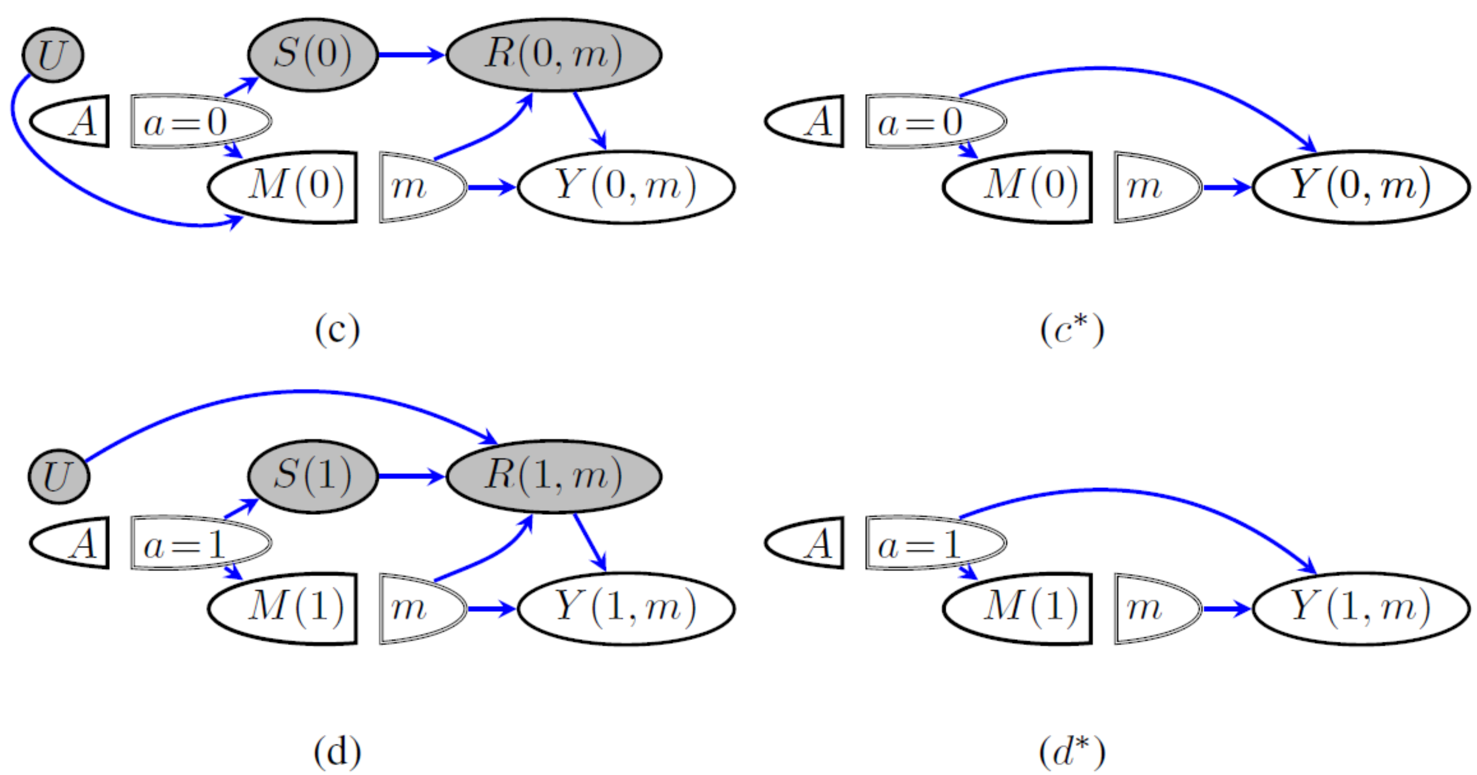
实例

- River Blindness Treatment Study
- A表示是否进入随机化实验，M表示是否接受药物，结局变量Y表示视力是否下降。S表示是否有机会得到抗过敏药，R表示是否接受抗过敏药。U为未观测混杂变量(是否接受药物治疗的倾向)。
- 由于某些原因，我们只有(A, M, Y)的数据。



CDE的识别性

- 在上述例子中，对A = 0的群体, U与M正相关. 而在A=1的群体中，由于M随机给定，所以U和M独立.
- $p(Y|A=a, M=m) = p(Y(a,m))$，从而CDE(m)可识别。

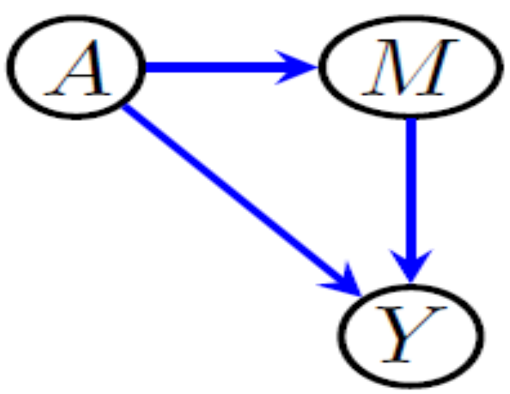


(c) (c*)

(d) (d*)

PDE的识别性

- 在上述例子中的PDE都是不识别的。
- 然后Robins讨论了更简单情形下的识别性。在该情形下，在NPSEM-IE(nonparametric structure equation model with independent error)模型中是可识别的
- 在FFRCISTG模型中不可识别，但可以做到部分识别(得到PDE的上下界)。



(a)

WHY?

$$\text{med}_{a,a'} \equiv \sum_m \left(\mathbb{E}[Y \mid m, a] - \mathbb{E}[Y \mid m, a']\right) p(m \mid a')$$
$$= \left(\sum_m \mathbb{E}[Y \mid m, a] p(m \mid a')\right) - \mathbb{E}[Y \mid a']$$

The proof of this result, under the NPSEM-IE is as follows:

$$\begin{aligned}
&p(Y(a, M(a') = m) = y) \\
&\quad = \sum_m p(Y(a, M(a') = m) = y \mid M(a') = m) p(M(a') = m) \\
&\quad = \sum_m p(Y(a, m) = y) p(M(a') = m) \\
&\quad = \sum_m p(Y = y \mid A = a, M = m) p(M = m \mid A = a')
\end{aligned}$$

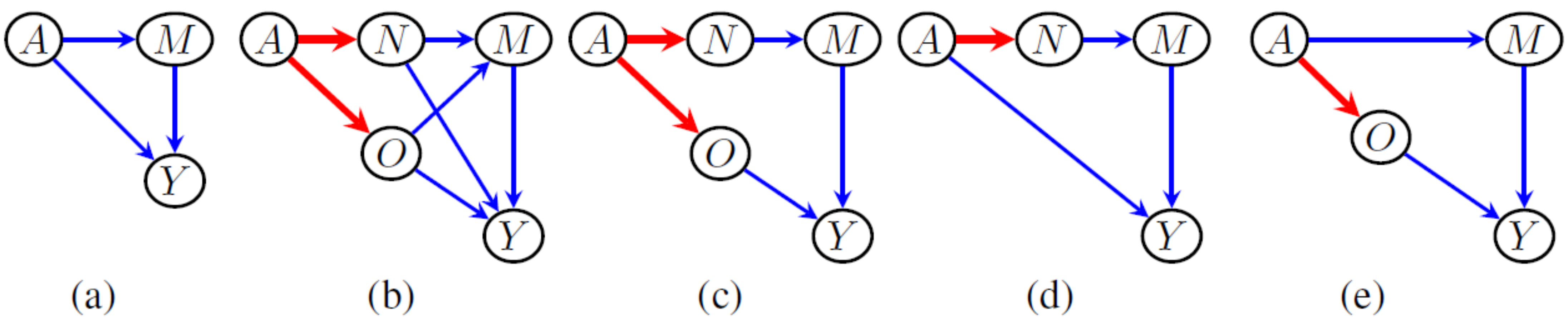
- This follows from the fact that the proof relies on the cross-world independence $Y(a, m) \perp M(a')$.

实例：Nicotine-Free Cigarette

- 有随机戒烟实验数据。吸烟状态A， 6个月后的高血压状态M，以及一年后的结局变量Y。
- 文章假设尼古丁对Y的作用都是通过M造成的，其他毒素对M没有作用。同时M到Y没有未观测的混杂变量。
- $PDE = E\left[Y\left(a = 1, M(a = 0)\right)\right] - E[Y(a = 0, M(a = 0))]$
- 而$\mathrm{E}[\mathrm{Y}(\mathrm{a} = 1, \mathrm{M}(\mathrm{a} = 0))] = \sum_m E[Y|A = 1, M = m]p(m|A = 0)$，从而PDE可识别。

Expanded DAG

Given a DAG $\mathcal{G}$ with a single treatment variable A, an *expanded graph* $\mathcal{G}^{ex}$ *for* A is a DAG constructed by first adding a set of new variables $\{A^{(1)}, \ldots, A^{(p)}\}$ corresponding to a decomposition of the treatment A into p separate components (proposed by the investigator); every variable $A^{(i)}$ is a child of A with the same state space and $A^{(i)}(a) = a$, but A has no other children in $\mathcal{G}^{ex}$; each child C_j of A in $\mathcal{G}$ has in $\mathcal{G}^{ex}$ a subset of $\{A^{(1)}, \ldots, A^{(p)}\}$ as its set of parents.



Expanded DAG

- $PDE = E[Y(n = 0, o = 1)] - E[Y(n = 0, o = 0)]$
- $PDE = \sum_m E[Y|O = 1, M = m]p(m|N = 0)$
- $\{O = 1, M = m\} = \{A = 1, M = m\}$
- $\{N = 0, M = m\} = \{A = 0, M = m\}$
- $PDE = \sum_m E[Y|A = 1, M = m]p(m|A = 0)$

Identification of four arms from two

- 文章给出了三种类型的数据
- 第一种：从实验中得到的原始数据(A,M,Y)，其中A是随机化的。
- 第二种：随机化(N,O)的得到的四组的数据，在每一组中(n,o)属于{0,1}^2。
- 第三种：第二种数据中n=o的这一部分数据。

Identification of four arms from two

- $p(M=m,Y=y|A=a)=$
  $p(M(n=a,o=a)=m,Y(n=a,o=a)=y|A=a)=$
  $p(M(n=a,o=a)=m,Y(n=a,o=a)=y)$
- 由上式可以看出第三种数据的分布可以由第一种数据识别。
- 从而我们的目标就是当$n\neq o$时去识别$E[Y(n,o)]$。

Identification of four arms from two

Proposition 1 *If for some $x \in \{0,1\}$ the following two conditions hold:*

$$p(M(n{=}x, o{=}0) = m) = p(M(n{=}x, o{=}1) = m), \tag{14}$$

$$\begin{aligned} &p(Y(n{=}1, o{=}x^*) = y \mid M(n{=}1, o{=}x^*) = m) \\ &\qquad = p(Y(n{=}0, o{=}x^*) = y \mid M(n{=}0, o{=}x^*) = m), \end{aligned} \tag{15}$$

where $x^* = 1 - x$, then:

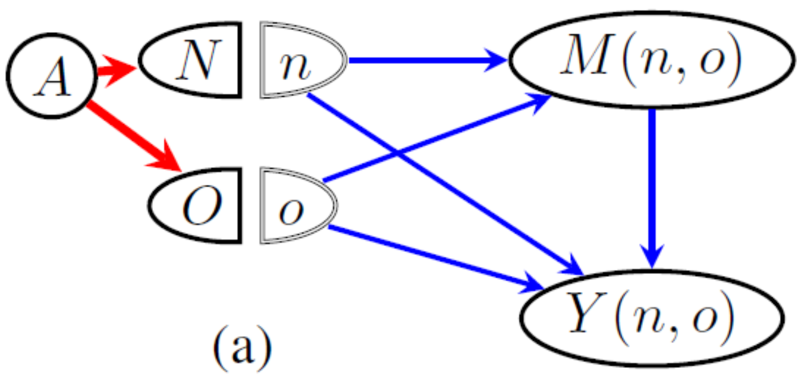
$$\begin{aligned} &p(M(n{=}x, o{=}x^*) = m, Y(n{=}x, o{=}x^*) = y) \\ &\quad = p(Y(n{=}x^*, o{=}x^*) = y \mid M(n{=}x^*, o{=}x^*) = m) p(M(n{=}x, o{=}x) = m). \end{aligned} \tag{16}$$

Proof:

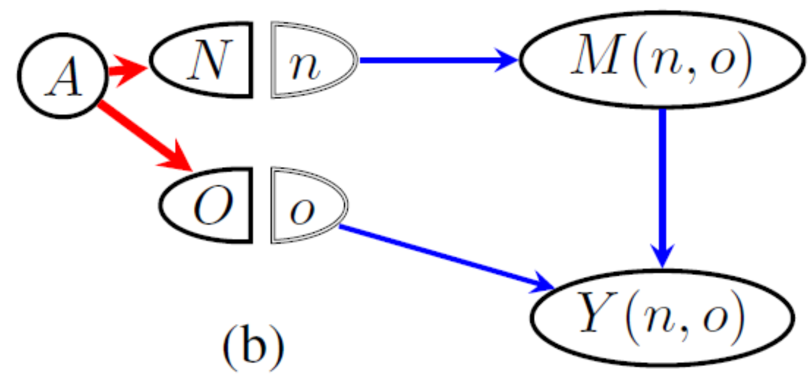
$$\begin{aligned} &p(M(n{=}x, o{=}x^*), Y(n{=}x, o{=}x^*)) \\ &\qquad = p(Y(n{=}x, o{=}x^*) \mid M(n{=}x, o{=}x^*)) p(M(n{=}x, o{=}x^*)) \\ &\qquad = p(Y(n{=}x^*, o{=}x^*) \mid M(n{=}x^*, o{=}x^*)) p(M(n{=}x, o{=}x)). \end{aligned}$$

Identification of four arms from two

Proposition 2 *Assume the distribution of the variables on an unexpanded DAG $\mathcal{G}$ is positive. Under an FFRCISTG corresponding to the population SWIG $\mathcal{G}^{ex}(n, o)$ there is no vertex that has both n and o as parents if and only if the joint distributions $p(V(n = x, o = x^*))$ for $x \neq x^*$ are identified from the counterfactual distributions $p(V(n=x, o=x))$. Further, since $P(V(n=x, o=x)) = P(V(a=x))$, also from the distribution of the variables in $\mathcal{G}$.*[30]



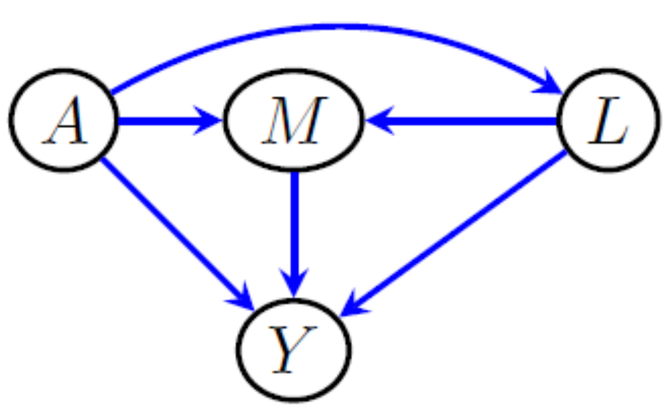
(a)



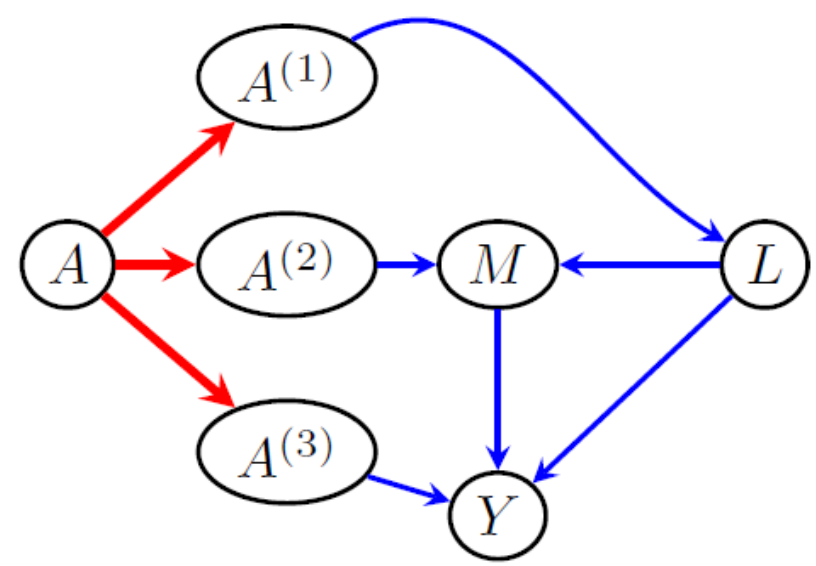
(b)

对于单个处理的识别性条件

Lemma 1 *Assume the distribution of the variables on an unexpanded DAG $\mathcal{G}$ is positive. Under an FFRCISTG corresponding to an expanded (population) graph $\mathcal{G}^{ex}$ for treatment A, the intervention distribution $p(V(a^{(1)} = x^{(1)}, \ldots, a^{(p)} = x^{(p)}))$ is identified by the g-formula applied to $\mathcal{G}^{ex}$ from the data on $\mathcal{G}$ if for every child C_j of A in $\mathcal{G}$, the set of parents of C_j in $\mathcal{G}^{ex}$ that are components of A take the same value.*[32]



(a)



Generalizations

Theorem 1 *If $V(\pi, a, a')$ is edge consistent, then under the NPSEM-IE for the DAG $\mathcal{G}$,*

$$p(V(\pi, a, a')) = \prod_{i=1}^{K} p(V_i \mid a \cap \mathrm{pa}_i^{\pi}, a' \cap \mathrm{pa}_i^{\overline{\pi}}, \mathrm{pa}_i^{\mathcal{G}} \setminus A). \tag{26}$$

Theorem 2 *Under the FFRCISTG model associated with the edge expanded DAG $\mathcal{G}^e$, for any edge consistent π, a, a':*

$$p^e(V(a^{\pi})) = \prod_{i=1}^{K} p^e(V_i \mid a^{\pi} \cap \mathrm{pa}_i^{\mathcal{G}^e}, \mathrm{pa}_i^{\mathcal{G}^e} \setminus A). \tag{27}$$

讨论

- 这篇文章将处理A分成了几个成分，然后去利用这成分来进行中介分析，并研究其识别性条件。
- 但是在实际应用时，A应该如何分解往往很困难。
- 同时FFRCISTG的相关论文很少，研究的人不多。

文献/资源列表

- James M. Robins. A new approach to causal inference in mortality studies with sustained exposure periods – application to control of the healthy worker survivor effect. Mathematical Modeling, 7:1393–1512, 1986.

- Richardson, Thomas S., and James M. Robins. "Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality." Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper 128.30 (2013): 2013.

- Richardson, Thomas S., and James M. Robins. "Single world intervention graphs: a primer." *Second UAI workshop on causal structure learning, Bellevue, Washington*. 2013.

联系我们

- 公众号：集智俱乐部
- 公众号：集智学园
- 添加读书会负责人为好友，参与线上讨论

集智学园

集智俱乐部

读书会负责人

更多"人工智能&复杂系统"的知识，请关注官网：pattern.swarma.org

集智斑图 Swarma Pattern