UNIVERSITÀ DI PAVIA

Probabilistic Graphical Models and Causal Inference

Episode 2: Structural Causal Models (from Interventions to Counterfactuals)

Marco Piastra

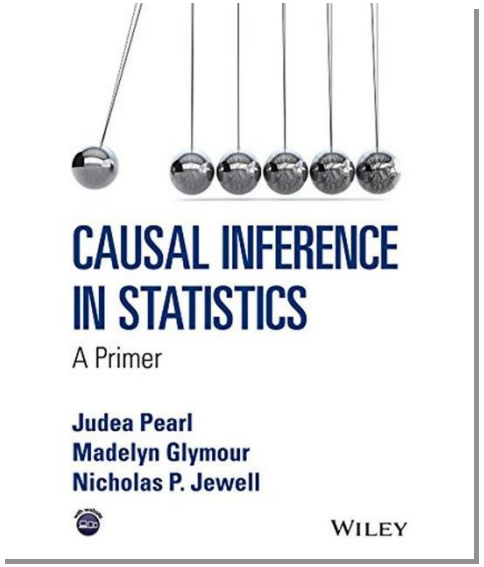
This presentation can be downloaded at:
https://vision.unipv.it/AI/AIRG.html

Very Good Readings

- **Causal Inference in Statistics**
  A Primer

  Judea Pearl, Madelyn Glymour and Nicholas P. Jewel

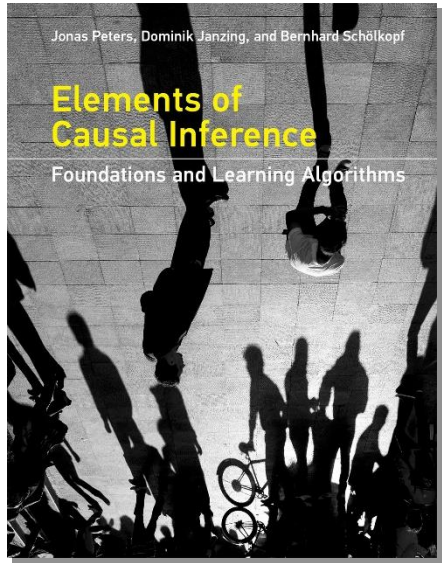
  Wiley, 2016



- **Elements of Causal Inference**
  Foundations and Learning Algorithms

  Jonas Peters, Dominik Janzing and Bernhard Schölkopf

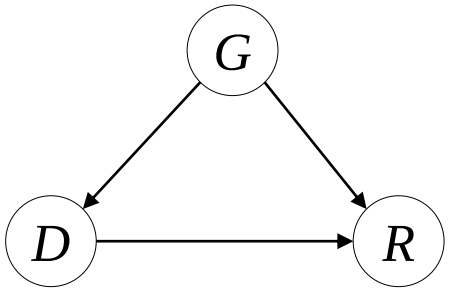
  MIT Press, 2017



Causal Graphical Models and Interventions (*recap*)

Causes and Effects: *say it with graphs*

- Probabilistic Graphical Model



G is *biological gender* (= Male/Female)
D is *drug administration* (= Yes(1)/No(0))
R is *recovery* from illness (= Yes(1)/No(0))

$$P(G, R, D) = P(G)P(D|G)P(R|G, D)$$

Females	$R = 0$	$R = 1$		Recovery Rate
$D = 0$	25	55	80	69%
$D = 1$	71	192	263	73%
	96	247	343	

Males	$R = 0$	$R = 1$		Recovery Rate
$D = 0$	36	234	270	87%
$D = 1$	6	81	87	93%
	42	315	357	

	$R = 0$	$R = 1$		Recovery Rate
$D = 0$	61	289	350	83%
$D = 1$	77	273	350	78%
	138	562	700	

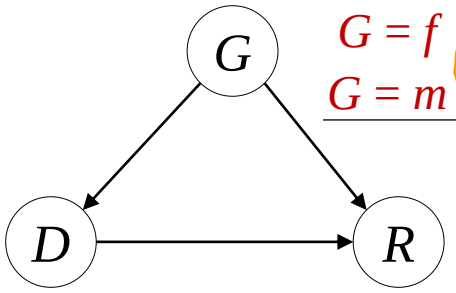
[Data from Pearl, J. et al., "Causal Inference in Statistics: A Primer", Wiley, 2016]

Causes and Effects: *say it with graphs*

- Probabilistic Graphical Model

  Maximum Likelihood Estimation (CPTs) of

$$P(G, R, D) = P(G)P(D|G)P(R|G, D)$$



$P(G)$

$G = f$	0.49
$G = m$	0.51

$P(D|G)$

	$G = f$	$G = m$
$D = 0$	0.23	0.76
$D = 1$	0.77	0.24

$P(R|G, D)$

	$G = f$ $D = 0$	$G = f$ $D = 1$	$G = m$ $D = 0$	$G = m$ $D = 1$
$R = 0$	0.31	0.27	0.13	0.07
$R = 1$	0.69	0.73	0.87	0.93

Females	$R = 0$	$R = 1$		Recovery Rate
$D = 0$	25	55	80	69%
$D = 1$	71	192	263	73%
	96	247	343	

Males	$R = 0$	$R = 1$		Recovery Rate
$D = 0$	36	234	270	87%
$D = 1$	6	81	87	93%
	42	315	357	

	$R = 0$	$R = 1$		Recovery Rate
$D = 0$	61	289	350	83%
$D = 1$	77	273	350	78%
	138	562	700	

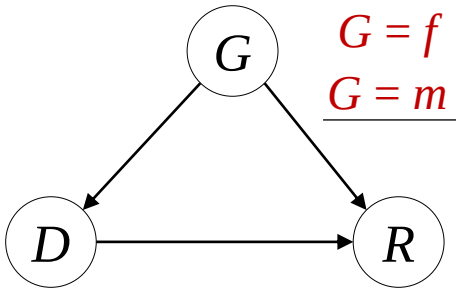
[Data from Pearl, J. et al., "Causal Inference in Statistics: A Primer", Wiley, 2016]

Causes and Effects: *say it with graphs*

- Probabilistic Graphical Model

  Maximum Likelihood Estimation (CPTs) of

$$P(G, R, D) = P(G)P(D|G)P(R|G, D)$$



$P(G)$

$G = f$	0.49
$G = m$	0.51

$P(D|G)$

	$G = f$	$G = m$
$D = 0$	0.23	0.76
$D = 1$	0.77	0.24

$P(R|G, D)$

	$G = f$ $D = 0$	$G = f$ $D = 1$	$G = m$ $D = 0$	$G = m$ $D = 1$
$R = 0$	0.31	0.27	0.13	0.07
$R = 1$	0.69	0.73	0.87	0.93

Using Graphical Model as a predictor

Case 1: Gender is observed

$$P(R = 1|G = 0, D = 0) = 0.69$$
$$P(R = 1|G = 0, D = 1) = 0.73$$
$$P(R = 1|G = 1, D = 0) = 0.87$$
$$P(R = 1|G = 1, D = 1) = 0.93$$

Prescribe drug, regardless

Case 2: Gender is <u>not</u> observed

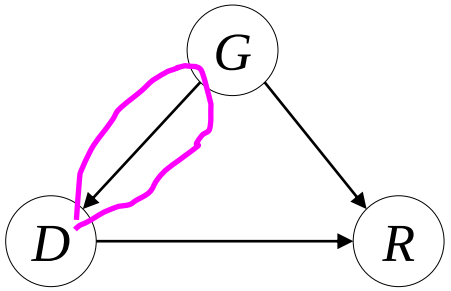
$$P(R|D) = \frac{\sum_G P(R|G, D)P(D|G)P(G)}{\sum_{G,R} P(R|G, D)P(D|G)P(G)}$$

$$P(R = 1|D = 0) = 0.83$$
$$P(R = 1|D = 1) = 0.78$$

Do not prescribe drug, regardless
(ridiculous!)

Causes and Effects: *say it with graphs*

- Probabilistic Graphical Model



G is *biological gender* (= Male/Female)
D is *drug administration* (= Yes(1)/No(0))
R is *recovery* from illness (= Yes(1)/No(0))

How can we solve the problem?

- The problem is due to the discrepancy in drug administration across genders
- An obvious solution would be *to repeat* the experiment with equal administration rates
- *In other words, we would sever this link*

Females	$R = 0$	$R = 1$		Recovery Rate
$D = 0$	25	55	80	69%
$D = 1$	71	192	263	73%
	96	247	343	

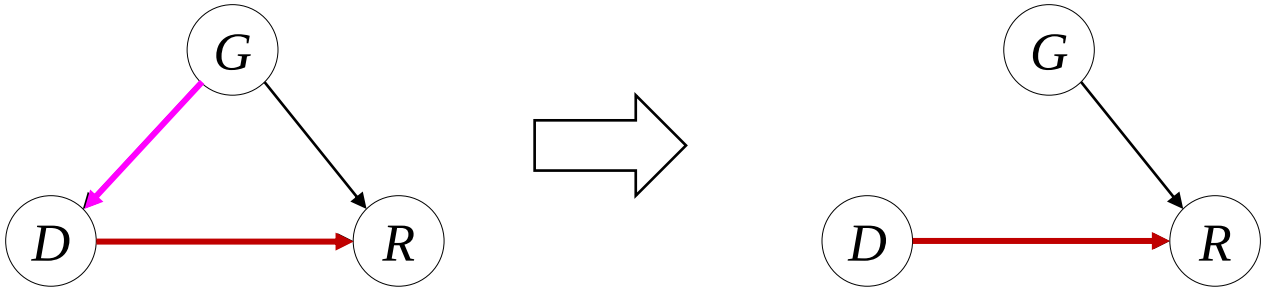
Males	$R = 0$	$R = 1$		Recovery Rate
$D = 0$	36	234	270	87%
$D = 1$	6	81	87	93%
	42	315	357	

	$R = 0$	$R = 1$		Recovery Rate
$D = 0$	61	289	350	83%
$D = 1$	77	273	350	78%
	138	562	700	

[Data from Pearl, J. et al., "Causal Inference in Statistics: A Primer", Wiley, 2016]

The Magic of Controlled Experiments

- **When association is causation**



In this *Causal Graphical Model*:

1. The causal effect we are interested in is that of D over R
2. The link between G and D is *problematic*: we know that $P(D|G=0) \neq P(D|G=1)$
3. In a *controlled experiment*, D is administered at random , therefore

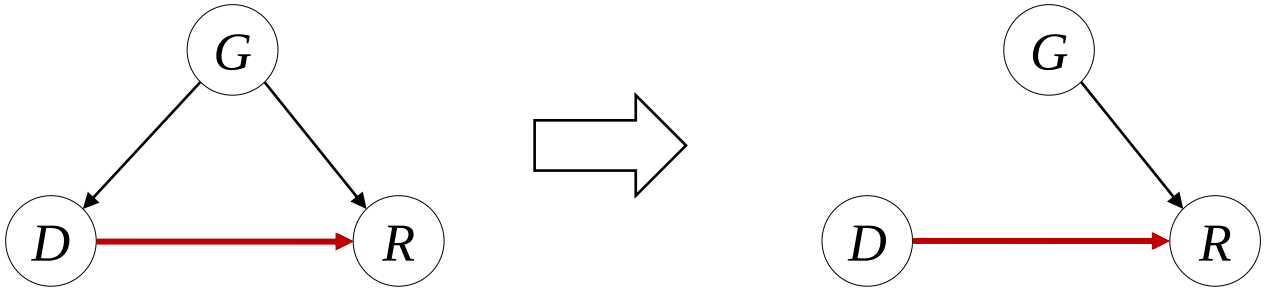
$$< D \perp G > \implies P(D|G=0) = P(D|G=1) = P(D)$$

4. In other words, the corresponding CGM 'loses' the problematic link and the estimate becomes

$$P(R|D) := \sum_G P(G)P(R|G,D)$$

The Magic of Controlled Experiments

- **When association is causation**

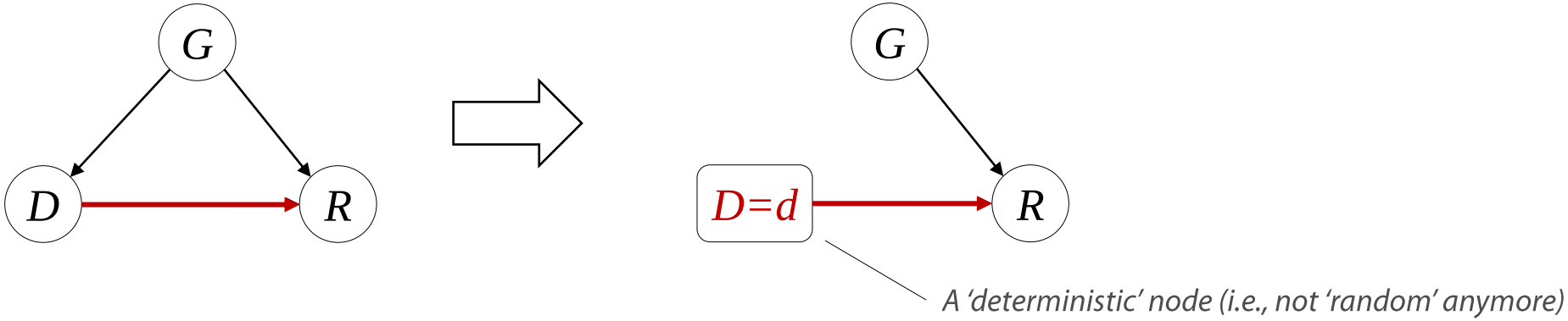


With *controlled experiments* (i.e., the 'gold standard' for testing) the principle is more general:

- by *randomizing* the administration of treatment
- we make the *effects* independent of any *confounders* (be them observed or not)

do-calculus

- **From Conditional (pre-intervention) to Intervention Probability**



In this *Causal Graphical Model* (for an <u>uncontrolled</u> experiment):

1. Conditional probability:

$$P(R|D=d) = \frac{\sum_G P(G)P(R|G,D=d)P(D=d|G)}{\sum_G P(G)P(D=d|G)}$$

2. Intervention (do-calculus, *this is new*)

$$P(R|do(D=d)) := \sum_G P(G)P(R|G,D=d)$$

These two expression would be identical if

$$P(D=d|G) = 1$$

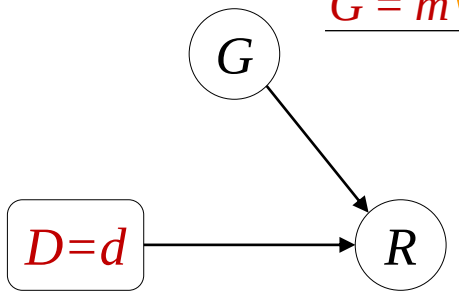
which cannot be true in general

3. *This is equivalent to* $P(R|D=d)$ *in a <u>modified</u> CGM in which we 'enforce intervention'*

do-calculus

- **From Conditional (pre-intervention) to Intervention Probability**

  (same observational probabilities, from MLE)



$P(G)$

$G = f$	0.49
$G = m$	0.51

$P(R|G, D)$

	$G = f$ $D = 0$	$G = f$ $D = 1$	$G = m$ $D = 0$	$G = m$ $D = 1$
$R = 0$	0.31	0.27	0.13	0.07
$R = 1$	0.69	0.73	0.87	0.93

Using do-calculus

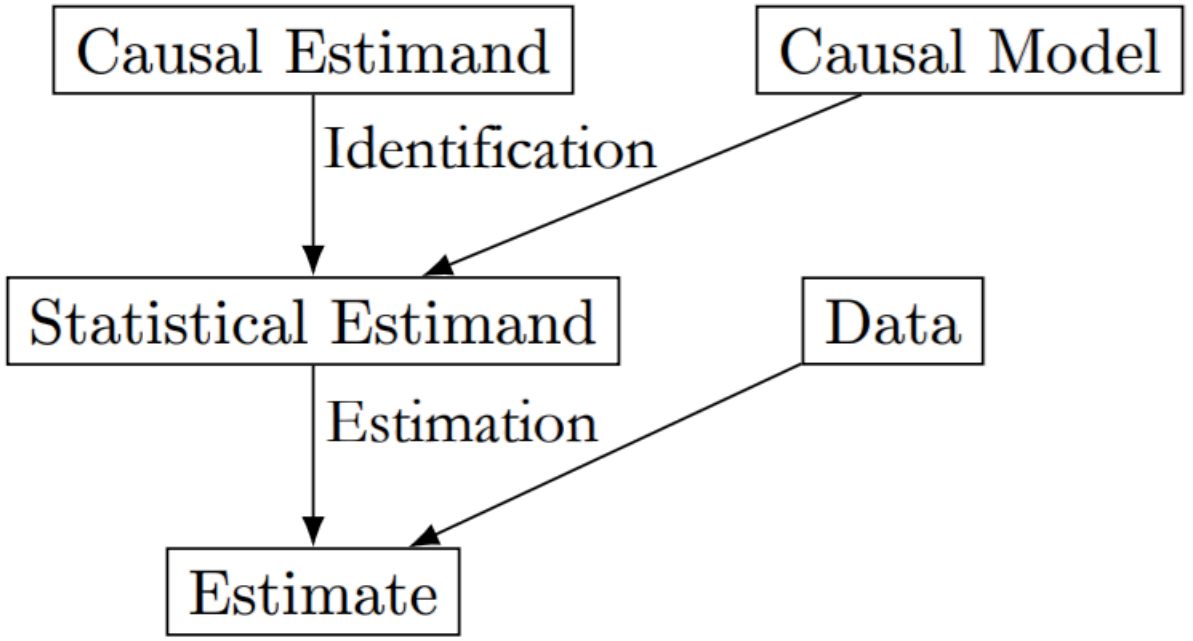
$$P(R = 1|do(D = 0)) = \sum_G P(G)P(R = 1|G, D = 0)$$
$$= 0.49 \cdot 0.69 + 0.51 \cdot 0.87 = 0.78$$

$$P(R = 1|do(D = 1)) = \sum_G P(G)P(R = 1|G, D = 1)$$
$$= 0.49 \cdot 0.73 + 0.51 \cdot 0.93 = 0.83$$

Prescribe drug, regardless

Causation and Conditionals

- **Causal Model and Estimation**



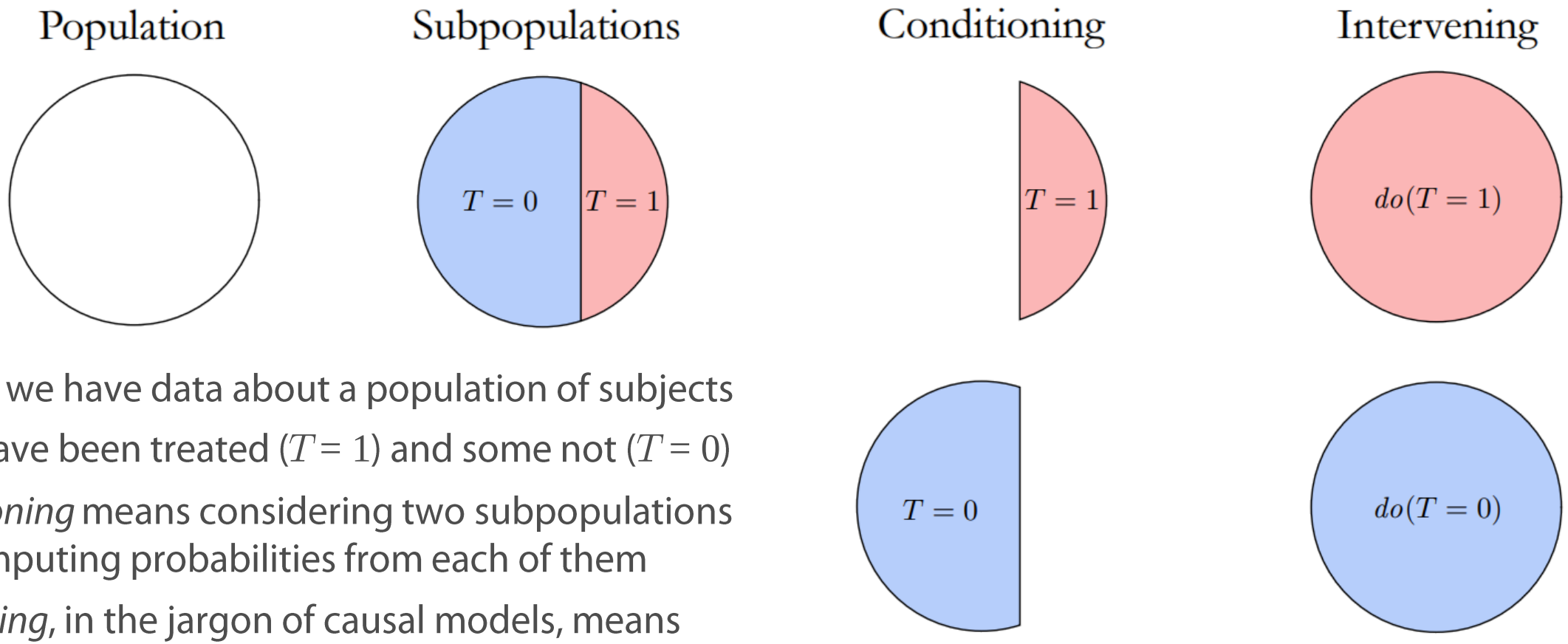
Basic principles:

1. Having selected what kind of causal effect we want to estimate
2. We start from a *Causal Graphical Model* (CGM)
3. To translate the estimate into a statistical estimand, (*Identification*)
4. We use then *observational* data to compute the estimate: a *probability* or an *expected value*

[Image from https://www.bradyneal.com/causal-inference-course]

Causation and Conditionals

- **Conditioning and Intervening**



Assume we have data about a population of subjects

Some have been treated ($T = 1$) and some not ($T = 0$)

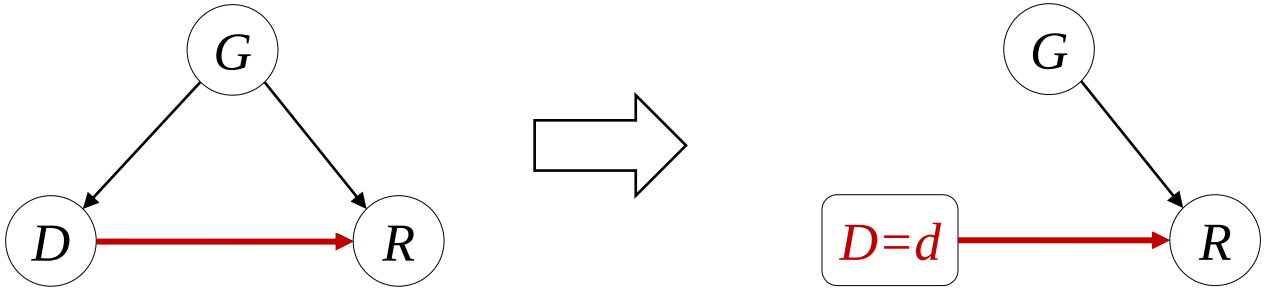
Conditioning means considering two subpopulations and computing probabilities from each of them

Intervening, in the jargon of causal models, means assuming that every subject in the population has been treated or not (*potential outcomes*)

[Image from https://www.bradyneal.com/causal-inference-course]

do-Calculus

- **Compare two expressions**



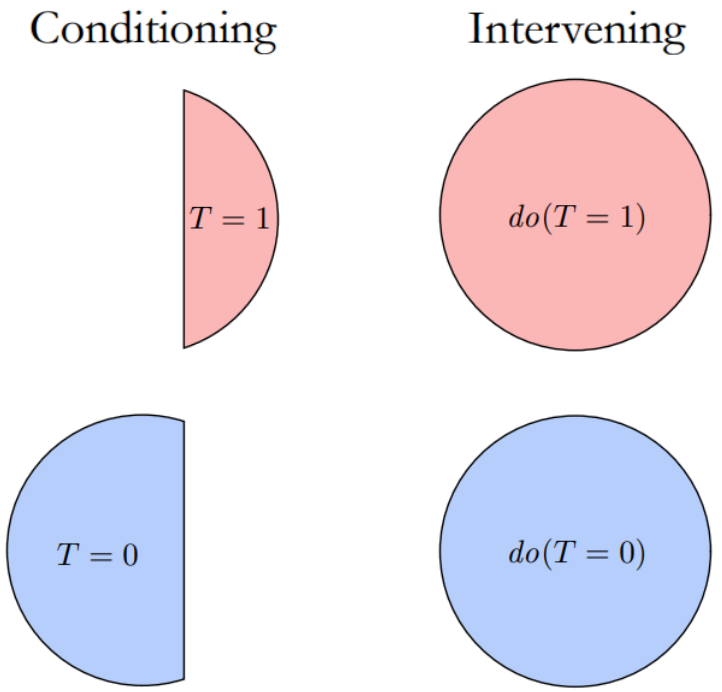
1. Conditioning:

$$P(R|D=d) = \frac{\sum_G P(G)P(R|G,D=d)P(D=d|G)}{\sum_G P(G)P(D=d|G)}$$

2. Intervening:

$$P(R|do(D=d)) := \sum_G P(G)P(R|G,D=d)$$

no normalization = no conditional subspace



Identification

- **Adjustment Set Criterion** [Shipster et al. 2010]

  In a Causal Graphical Model, the *causal effect* of T over Y is *identifiable*
  iff it exists an *adjustment set* $\boldsymbol{W}$ of variables such that:

  - no variable in $\boldsymbol{W}$ is on, or is a descendant of any variables on, a causal path
    (excluding the descendants of T alone)
  - the variables in $\boldsymbol{W}$ block *(in the sense of graphical models)*
    all the non-causal paths between T and Y

  This criterion is necessary and sufficient for effect identifiability

  Then:

$$P(Y|do(T=t)) = \sum_{\boldsymbol{W}} P(Y|T=t,\boldsymbol{W})P(\boldsymbol{W})$$

  In words, the causal effect can be estimated statistically, from data

Identification

- **Adjustment Set Criterion with *observed* and *unobserved* variables**

  More in general, in practical cases,
  there can be both observed and unobserved (possibly hidden) variables

  An *adjustment set* can be composed of both:

  $$\boldsymbol{W} = \boldsymbol{W}_{obs} \cup \boldsymbol{W}_{hid}$$

  Then, if $\boldsymbol{W}$ satisfies the Adjustment Set Criterion:

  $$P(Y|do(T=t), \boldsymbol{W}_{obs}) = \sum_{\boldsymbol{W}_{hid}} P(Y|T=t, \boldsymbol{W}_{hid}, \boldsymbol{W}_{obs}) P(\boldsymbol{W}_{hid})$$

  When there are no *observed* variables in the adjustment set:

  $$P(Y|do(T=t)) = \sum_{\boldsymbol{W}} P(Y|T=t, \boldsymbol{W}) P(\boldsymbol{W})$$

  Likewise, when there are no *unobserved* variables in the adjustment set:

  $$P(Y|do(T=t), \boldsymbol{W}) = P(Y|T=t, \boldsymbol{W})$$

Estimating Effects

Expected effects of different interventions can be estimated via do-calculus

In general, the *expected effect* on Y of treatment T will be

$$\mathbb{E}[Y|T=t, \boldsymbol{W}_{obs}] := \sum_{y \in \mathcal{Y}} y \; P(Y|do(T=t), \boldsymbol{W}_{obs})$$

where $\boldsymbol{W} = \boldsymbol{W}_{obs} \cup \boldsymbol{W}_{hid}$ is a valid *adjustment set*

Differences in effects can be measured by comparing expected effects.

As a special case, when $T \in \{0, 1\}$

- The *Conditional Average Treatment Effect* (CATE) is defined as:

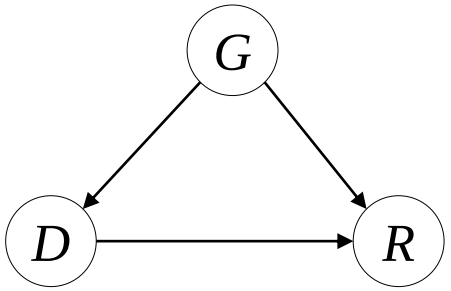
$$\tau(\boldsymbol{W}_{obs}) := \; \mathbb{E}[Y|T=1, \boldsymbol{W}_{obs}] - \mathbb{E}[Y|T=0, \boldsymbol{W}_{obs}]$$

- The *Average Treatment Effect* (ATE) is defined as:

$$\mathbb{E}[\tau(\boldsymbol{W})] := \; \mathbb{E}[Y|T=1] - \mathbb{E}[Y|T=0]$$

Structural Causal Models

Probabilistic Graphical Model



G is *biological gender* (= Male/Female)
D is *drug administration* (= Yes(1)/No(0))
R is *recovery* from illness (= Yes(1)/No(0))

$$P(G, R, D) = P(G)P(D|G)P(R|G, D)$$

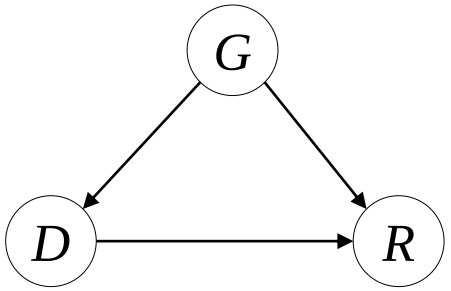
Females	$R = 0$	$R = 1$		Recovery Rate
$D = 0$	25	55	80	69%
$D = 1$	71	192	263	73%
	96	247	343	

Males	$R = 0$	$R = 1$		Recovery Rate
$D = 0$	36	234	270	87%
$D = 1$	6	81	87	93%
	42	315	357	

	$R = 0$	$R = 1$		Recovery Rate
$D = 0$	61	289	350	83%
$D = 1$	77	273	350	78%
	138	562	700	

[Data from Pearl, J. et al., "Causal Inference in Statistics: A Primer", Wiley, 2016]

From Graphical Model to Structural Equations



G is *biological gender* (= Male/Female)
D is *drug administration* (= Yes(1)/No(0))
R is *recovery* from illness (= Yes(1)/No(0))

$$P(G, R, D) = P(G)P(D|G)P(R|G, D)$$

** first approximation*

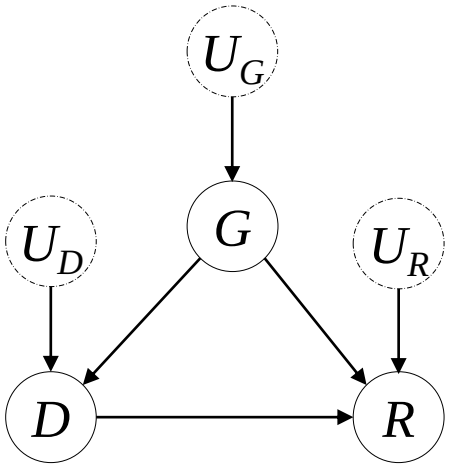
Structural Equations

$$D = f_D(G)$$

$$R = f_R(G, D)$$

How can these two things be reconciled?
Functions are deterministic

From Graphical Model to Structural Equations



** second approximation*

Structural Equations

$$G = U_G$$

$$D = f_D(G, U_D)$$

$$R = f_R(G, D, U_R)$$

G is *biological gender* (= Male/Female)
D is *drug administration* (= Yes(1)/No(0))
R is *recovery* from illness (= Yes(1)/No(0))

$$P(G, R, D) = P(G)P(D|G)P(R|G, D)$$

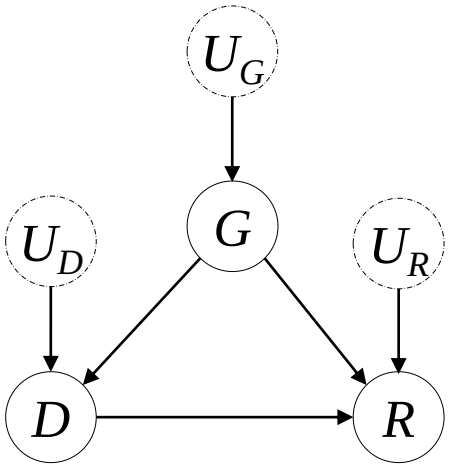
Causal? Functions could be invertible
Example:

$$D = k + \beta_g G + U_D$$

$$G = \frac{1}{\beta_g}(D - k - U_D)$$

U_G, U_D and U_R are <u>unobservable</u>, <u>random</u> variables
The probability distribution is the <u>observable</u> aspect of the structural equations

From Graphical Model to Structural Causal Model



G is *biological gender* (= Male/Female)
D is *drug administration* (= Yes(1)/No(0))
R is *recovery* from illness (= Yes(1)/No(0))

$$P(G, R, D) = P(G)P(D|G)P(R|G, D)$$

Structural Causal Model (SCM)

$$G := U_G$$

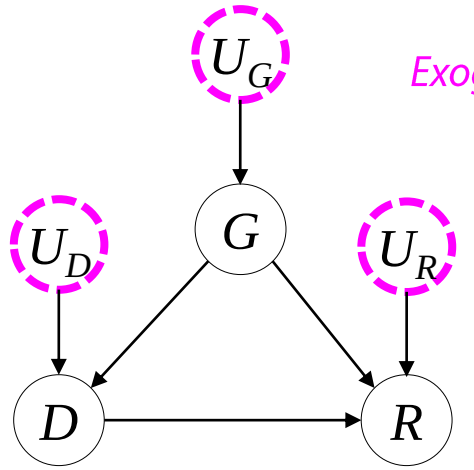
$$D := f_D(G, U_D)$$

$$R := f_R(G, D, U_R)$$

Force directions, in keeping with causation assumptions

U_G, U_D and U_R are unobservable, random variables
The probability distribution is the observable aspect of the structural causal model

From Graphical Model to Structural Causal Model



Exogenous variables

Structural Causal Model (SCM)

$$G := U_G$$

$$D := f_D(G, U_D)$$

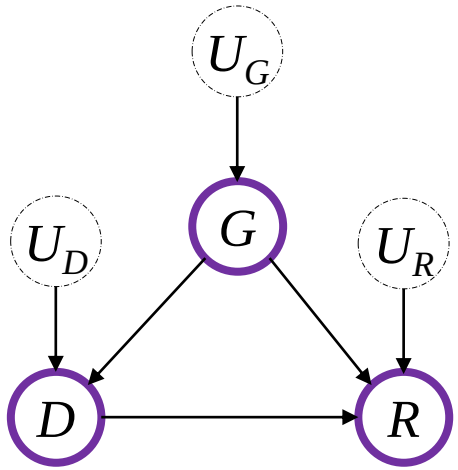
$$R := f_R(G, D, U_R)$$

G is *biological gender* (= Male/Female)
D is *drug administration* (= Yes(1)/No(0))
R is *recovery* from illness (= Yes(1)/No(0))

$$P(G, R, D) = P(G)P(D|G)P(R|G, D)$$

U_G, U_D and U_R are unobservable, random variables
The probability distribution is the observable aspect of the structural causal model

From Graphical Model to Structural Causal Model



Endogenous variables

Structural Causal Model (SCM)

$$G := U_G$$

$$D := f_D(G, U_D)$$

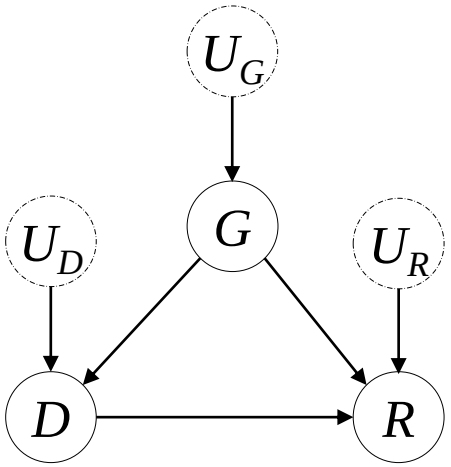
$$R := f_R(G, D, U_R)$$

G is *biological gender* (= Male/Female)
D is *drug administration* (= Yes(1)/No(0))
R is *recovery* from illness (= Yes(1)/No(0))

$$P(G, R, D) = P(G)P(D|G)P(R|G, D)$$

U_G, U_D and U_R are unobservable, random variables
The probability distribution is the observable aspect of the structural causal model

Structural Causal Model



Structural Causal Model (SCM)

$$G := U_G$$

$$D := f_D(G, U_D)$$

$$R := f_R(G, D, U_R)$$

Structural Causal Model (SCM) definition

1) A set of *endogenous* variables
2) A set of *exogenous* variables
3) A set of *structural equations*

An SCM *induces* a graphical model with a probability distribution P over *endogenous* variables

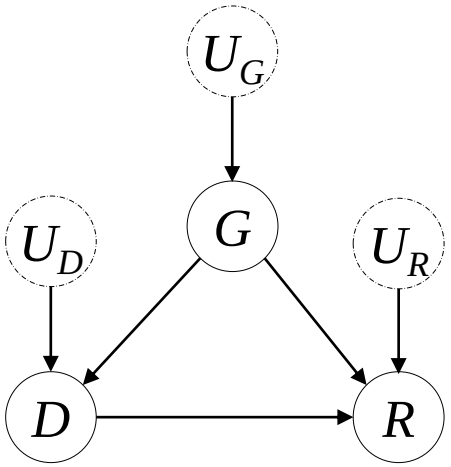
Structural Causal Model (formal definition)

Structural Causal Model (SCM), formally

1) A set of *endogenous* variables $\boldsymbol{V} := \{V_1, V_2, \cdots, V_n\}$
2) A set of *exogenous* variables $\boldsymbol{U} := \{U_1, U_2, \cdots, U_n\}$
3) A set of *structural equations* $\boldsymbol{f} := \{f_1, f_2, \cdots, f_n\}$

An SCM $\mathcal{M}$ *induces* a graphical model $\mathcal{G}$ with a probability distribution $P(\boldsymbol{V})$

Structural Causal Model



induces

Structural Causal Model (SCM)

$$G := U_G$$

$$D := f_D(G, U_D)$$

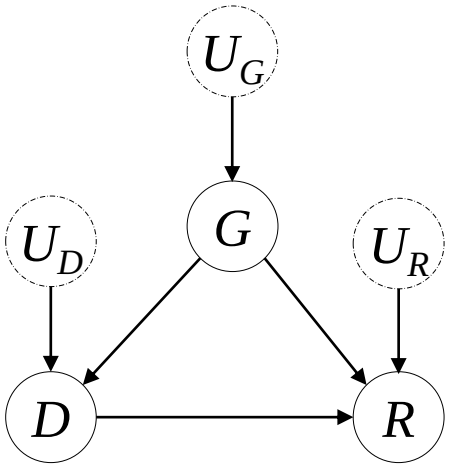
$$R := f_R(G, D, U_R)$$

The graphical model induced is uniquely defined

Further questions:

1) *Which functions?*
2) *How are the random variables U_G, U_D and U_R distributed?*
3) *Are they dependent (or correlated)?*
4) *Is the SCM* identifiable *from observed data?*

Structural Equation Model (linear functions)



a special case, linear

Structural Equation Model (SEM)

$$G := U_G$$

$$D := k_1 + \beta_1 G + U_D$$

$$R := k_2 + \beta_2 G + \beta_3 D + U_R$$

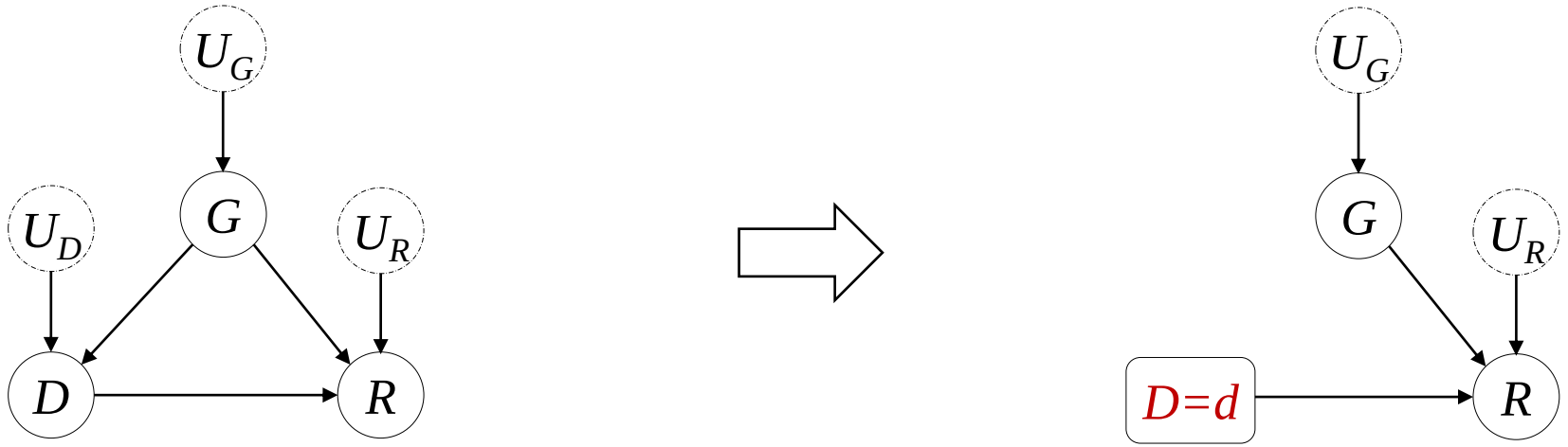
Assumptions:

1) *All functions are linear*
2) *All random variables U_G, U_D and U_R are normally distributed*
3) *All random variables are uncorrelated*

Under further, specific conditions a SEM is identifiable from observed data

more in general, however, this is not true of any SCM

Intervention in a Structural Causal Model



Structural Equation Model (SEM)

$$G := U_G$$
$$D := f_D(G, U_D)$$
$$R := f_R(G, D, U_R)$$

$$G := U_G$$
$$D := d$$
$$R := f_R(G, D, U_R)$$

An intervention on an SCM creates a new sub-model by changing one or more structural equations
It induces a new graph

Counterfactuals

Counterfactuals?

The ladder of causal inference [J. Pearl, *Causation*, Cambridge University Press, 2009]

- **Prediction**

  Given the probability distribution $P(\boldsymbol{V})$ *and some observations* $\boldsymbol{V}_o = \boldsymbol{v}_o$
  determine the probability $P(\boldsymbol{V}_u = \boldsymbol{v}_u \mid \boldsymbol{V}_o = \boldsymbol{v}_o)$ *for some unobserved variables* $\boldsymbol{V}_u$

- **Intervention**

  Intervene (i.e., force a change in value) on some variables $\boldsymbol{V}_i$ *and determine*
  the probability of effects $P(\boldsymbol{V}_e = \boldsymbol{v}_e \mid do(\boldsymbol{V}_i = \boldsymbol{v}_i))$

- **Counterfactual**

  Having observed $\boldsymbol{V}_o = \boldsymbol{v}_o$ *and its effects* $\boldsymbol{V}_e = \boldsymbol{v}_e$ *, what could be the probability*
  of different effects $\boldsymbol{v}'_e \neq \boldsymbol{v}_e$ *if some conditions* $\boldsymbol{V}_c \subseteq \boldsymbol{V}_o$ *were different?*

Counterfactual Inference

- **Counterfactual**

  Having observed $\boldsymbol{V}_o = \boldsymbol{v}_o$ *and its effects* $\boldsymbol{V}_e = \boldsymbol{v}_e$ *, what could be the probability of different effects* $\boldsymbol{v}'_e \neq \boldsymbol{v}_e$ *if some conditions* $\boldsymbol{V}_c \subseteq \boldsymbol{V}_o$ *were different?*

  A few relevant aspects:

  - *Prediction and Intervention occur in the same world, whereas counterfactuals require alternative worlds*
  - *Conceptually, counterfactuals relate to potential outcomes ("what could it be the outcome, were the condition different?")*
  - *Counterfactual inference can be performed at either individual or population level (more to follow)*

Counterfactual Inference

- **General Method**

  A few relevant aspects:

  - *Prediction and Intervention occur in the same world, whereas counterfactuals require alternative worlds*
  - *Conceptually, counterfactuals relate to potential outcomes ("what could it be the outcome, were the condition different?")*
  - *Counterfactual inference can be performed at either individual or population level (more to follow)*

Counterfactual Inference, deterministic case

From J. Pearl, *Primer*, 2016 *(see suggested readings)*

1. Abduction *(i.e., going in reverse)*: use a specific observation $\boldsymbol{v}_o$ to determine the values $\boldsymbol{u}$ of the unobservable variables $\boldsymbol{U}$
2. Action: *create a sub-model of* $\mathcal{M}$ *by replacing the structural equations for* $\boldsymbol{V}_c$ *with* $\boldsymbol{V}_c = \boldsymbol{v}_c$ *(counterfactual values)*
3. Prediction: *use the sub-model to compute* $\boldsymbol{v}_e$ *(the effects) by using* $\boldsymbol{u}$, $\boldsymbol{v}_c$ *and* $\boldsymbol{v}_o \setminus \boldsymbol{v}_c$

all remaining observed values

Counterfactual Inference, deterministic case

From J. Pearl, *Primer*, 2016 *(see suggested readings)*

1. Abduction *(i.e., going in reverse)*: use a specific observation $\boldsymbol{v}_o$ to determine the values $\boldsymbol{u}$ of the unobservable variables $\boldsymbol{U}$
2. Action: *create a sub-model of* $\mathcal{M}$ *by replacing the structural equations for* $\boldsymbol{V}_c$ *with* $\boldsymbol{V}_c = \boldsymbol{v}_c$ *(counterfactual values)*
3. Prediction: *use the sub-model to compute* $\boldsymbol{v}_e$ *(the effects) by using* $\boldsymbol{u}$, $\boldsymbol{v}_c$ *and* $\boldsymbol{v}_o \setminus \boldsymbol{v}_c$

Example, in the linear case:

$$G := U_G$$

$$D := k_1 + \beta_1 G + U_D$$

$$R := k_2 + \beta_2 G + \beta_3 D + U_R$$

Assume all parameters are known (complete identification of $\mathcal{M}$ *)*

Counterfactual Inference, deterministic case

From J. Pearl, *Primer*, 2016 *(see suggested readings)*

1. Abduction *(i.e., going in reverse)*: use a specific observation $\boldsymbol{v}_o$ to determine the values $\boldsymbol{u}$ of the unobservable variables $\boldsymbol{U}$
2. Action: *create a sub-model of* $\mathcal{M}$ *by replacing the structural equations for* $\boldsymbol{V}_c$ *with* $\boldsymbol{V}_c = \boldsymbol{v}_c$ *(counterfactual values)*
3. Prediction: *use the sub-model to compute* $\boldsymbol{v}_e$ *(the effects) by using* $\boldsymbol{u}$, $\boldsymbol{v}_c$ *and* $\boldsymbol{v}_o \setminus \boldsymbol{v}_c$

Example, in the linear case:

$$u_G = g_o$$

$$u_D := d_o - k_1 - \beta_1 g_o$$

$$u_R := r_o - k_2 - \beta_2 g_o + \beta_3 d_o$$

Replace with observed values and solve for $\boldsymbol{U}$

Counterfactual Inference, deterministic case

From J. Pearl, *Primer*, 2016 *(see suggested readings)*

1. Abduction *(i.e., going in reverse)*: use a specific observation $\boldsymbol{v}_o$ to determine the values $\boldsymbol{u}$ of the unobservable variables $\boldsymbol{U}$
2. Action: *create a sub-model of* $\mathcal{M}$ *by replacing the structural equations for* $\boldsymbol{V}_c$ *with* $\boldsymbol{V}_c = \boldsymbol{v}_c$ *(counterfactual values)*
3. Prediction: *use the sub-model to compute* $\boldsymbol{v}_e$ *(the effects) by using* $\boldsymbol{u}$, $\boldsymbol{v}_c$ *and* $\boldsymbol{v}_o \setminus \boldsymbol{v}_c$

Example, in the linear case:

$$g = u_G$$

$$d = \boxed{d_c} \text{ counterfactual value}$$

$$\text{effect } \boxed{r_e} = k_2 + \beta_2 g_o + \beta_3 d_c + u_R$$

Plug back values $\boldsymbol{u}$, *impose counterfactual value* d_c *and compute the resulting effect* r_e

Counterfactual Inference, probabilistic case

From J. Pearl, *Primer*, 2016 *(see suggested readings)*

1. Abduction *(i.e., going in reverse)*: use a specific observation $\boldsymbol{v}_o$ to determine the values $\boldsymbol{u}$ of the unobservable variables $\boldsymbol{U}$
2. Action: *create a sub-model of* $\mathcal{M}$ *by replacing the structural equations for* $\boldsymbol{V}_c$ *with* $\boldsymbol{V}_c = \boldsymbol{v}_c$ *(counterfactual values)*
3. Prediction: *use the sub-model to compute* $\boldsymbol{v}_e$ *(the effects) by using* $\boldsymbol{u}$, $\boldsymbol{v}_c$ *and* $\boldsymbol{v}_o \setminus \boldsymbol{v}_c$

More in general, even keeping the assumption of complete identification of $\mathcal{M}$, *what happens if some functions are not one-to-one for the values of* $\boldsymbol{U}$ *?*

Counterfactual Inference, probabilistic case

From J. Pearl, *Primer*, 2016 *(see suggested readings)*

1. Abduction *(i.e., going in reverse)*: use a specific observation $\boldsymbol{v}_o$ to determine the values $\boldsymbol{u}$ of the unobservable variables $\boldsymbol{U}$
2. Action: *create a sub-model of* $\mathcal{M}$ *by replacing the structural equations for* $\boldsymbol{V}_c$ *with* $\boldsymbol{V}_c = \boldsymbol{v}_c$ *(counterfactual values)*
3. Prediction: *use the sub-model to compute* $\boldsymbol{v}_e$ *(the effects) by using* $\boldsymbol{u}$, $\boldsymbol{v}_c$ *and* $\boldsymbol{v}_o \setminus \boldsymbol{v}_c$

Example, non-invertible case:

$$G := U_G$$

$$D := f_D(G, U_D)$$

$$R := f_R(G, D, U_R)$$

$\mathcal{M}$ *is still completely identified*

Counterfactual Inference, probabilistic case

From J. Pearl, *Primer*, 2016 *(see suggested readings)*

1. Abduction *(i.e., going in reverse)*: use a specific observation $\boldsymbol{v}_o$ to determine the values $\boldsymbol{u}$ of the unobservable variables $\boldsymbol{U}$
2. Action: *create a sub-model of* $\mathcal{M}$ *by replacing the structural equations for* $\boldsymbol{V}_c$ *with* $\boldsymbol{V}_c = \boldsymbol{v}_c$ *(counterfactual values)*
3. Prediction: *use the sub-model to compute* $\boldsymbol{v}_e$ *(the effects) by using* $\boldsymbol{u}$, $\boldsymbol{v}_c$ *and* $\boldsymbol{v}_o \setminus \boldsymbol{v}_c$

Example, non-invertible case:

$$G = g_o$$

$$D = d_c$$

there might exist multiple values u_R *compatible with the observed effect* r_o

$$R_e = f_R(g_o, d_c, U_R)$$

$\mathcal{M}$ *is still completely identified*

Counterfactual Inference, probabilistic case

From J. Pearl, *Primer*, 2016 *(see suggested readings)*

1. Abduction *(i.e., going in reverse)*: use a specific observation $\boldsymbol{v}_o$ to update the probability distribution $P(\boldsymbol{U}|\boldsymbol{V}_o = \boldsymbol{v}_o)$
2. Action: *create a sub-model of* $\mathcal{M}$ *by replacing the structural equations for* $\boldsymbol{V}_c$ *with* $\boldsymbol{V}_c = \boldsymbol{v}_c$ *(counterfactual values)*
3. Prediction: *use the sub-model and the updated distribution to compute the probability of each possible effect* $\boldsymbol{v}_e$

Note that these two may be in contrast with each other

Counterfactual Inference, probabilistic case

From J. Pearl, *Primer*, 2016 *(see suggested readings)*

1. Abduction *(i.e., going in reverse)*: use a specific observation $\boldsymbol{v}_o$ to update the probability distribution $P(\boldsymbol{U}|\boldsymbol{V}_o = \boldsymbol{v}_o)$
2. Action: *create a sub-model of* $\mathcal{M}$ *by replacing the structural equations for* $\boldsymbol{V}_c$ *with* $\boldsymbol{V}_c = \boldsymbol{v}_c$ *(counterfactual values)*
3. Prediction: *use the sub-model and the updated distribution to compute the probability of each possible effect* $\boldsymbol{v}_e$

Example, non-invertible case:

$$G = g_o$$

$$D = d_c$$

$$P(R_e = r_e) = \sum_{\{u_R \mid f_R(g_o, d_c, u_R) = r_e\}} P(u_R|\boldsymbol{V}_o = \boldsymbol{v}_o)$$

all the values u_R *compatible with the effect* r_e

Counterfactual Inference, probabilistic case

From J. Pearl, *Primer*, 2016 *(see suggested readings)*

1. Abduction *(i.e., going in reverse)*: use a specific observation $\boldsymbol{v}_o$ to update the probability distribution $P(\boldsymbol{U}|\boldsymbol{V}_o = \boldsymbol{v}_o)$
2. Action: *create a sub-model of* $\mathcal{M}$ *by replacing the structural equations for* $\boldsymbol{V}_c$ *with* $\boldsymbol{V}_c = \boldsymbol{v}_c$ *(counterfactual values)*
3. Prediction: *use the sub-model and the updated distribution to compute the probability of each possible effect* $\boldsymbol{v}_e$

This is called unit or individual-level counterfactual inference since it starts from the observation (possibly complete) of a specific case

It requires the complete identification of $\mathcal{M}$ *(including the distribution* $P(\boldsymbol{U}, \boldsymbol{V})$ *)*

Otherwise, there are too many degrees of freedom and the inference problem is ill-posed

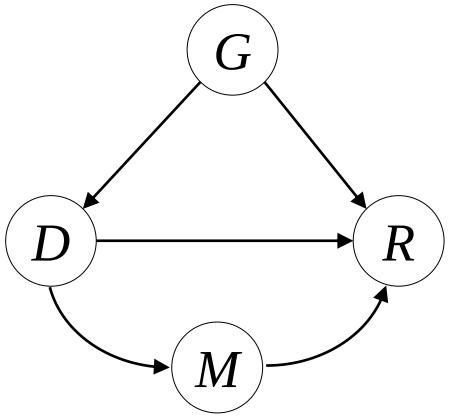
Counterfactual Inference, population level

What kind of counterfactual inference can be performed when the model $\mathcal{M}$ is NOT completely identified?

In other words, when what we have is the distribution $P(\boldsymbol{V})$ over endogenous variables as derived from actual observations?

Perhaps we should change the question somewhat:
what other kind of counterfactual inference could be useful in such case?

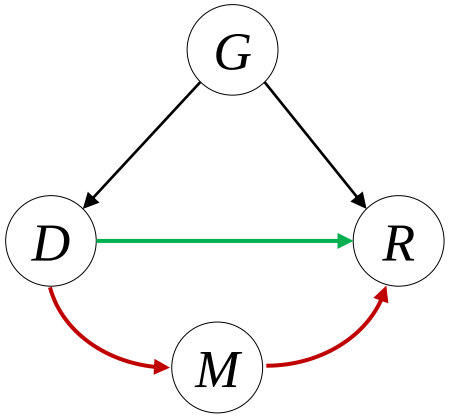
Path-Specific Counterfactual Evidence



Suppose that, as an extension to the previous model, we now assume that drug D has an observable side-effect M which also affects patient's recovery R

It is independent from gender G

Path-Specific Counterfactual Evidence



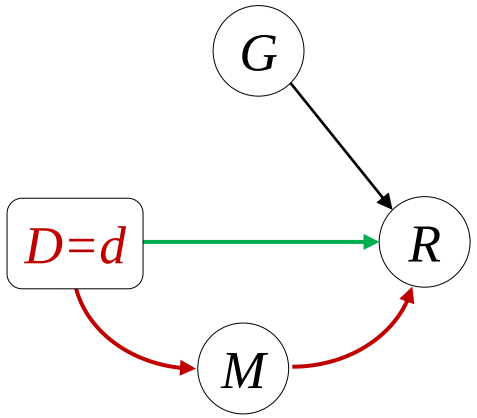
Suppose that, as an extension to the previous model, we now assume that drug D has an observable side-effect M which also affects patient's recovery R

It is independent from gender G

Now the model has two causal paths: one direct and another indirect, mediated by M

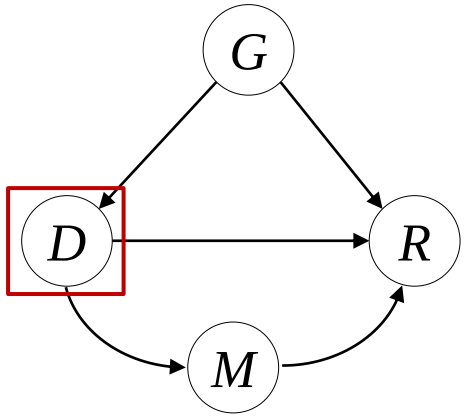
We might want to know what are the causal effects of each path in general, i.e., at the population level

Path-Specific Counterfactual Evidence



Intervention on D alone will not give the answer, as both paths need to be considered at once

Path-Specific Counterfactual Evidence

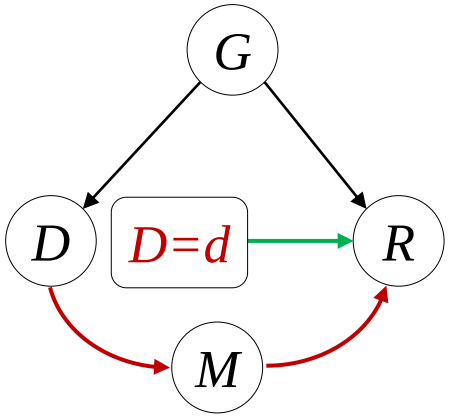


General idea (intuitive): splitting node D in two and letting different paths 'see' different values

SCM model $\mathcal{M}$

$$G := U_G$$
$$D := f_D(G, U_D)$$
$$M := f_M(D, U_M)$$
$$R := f_R(G, D, M, U_R)$$

Path-Specific Counterfactual Evidence



General idea (intuitive): splitting node D in two and letting different paths 'see' different values

Modified SCM model $\mathcal{M}'$

$$G := U_G$$
$$D := f_D(G, U_D)$$
$$M := f_M(D, U_M)$$
$$R := f_R(G, \boxed{D = d}, M, U_R)$$

When this is feasible, differences in path-specific effects can be evaluated from the distribution $P(\boldsymbol{V})$ *alone*

Counterfactual Inference
(see GeNIe 'berkeley_path_specific' attachment)

Identifiability of Path-Specific Effects

- **Path-Specific Criterion (simplified)**

  In a SCM model $\mathcal{M}$, *path-specific effects* of T over Y with mediator M are *identifiable* iff it exists an *adjustment set* $\boldsymbol{W}$ of variables such that:

  - no variable in $\boldsymbol{W}$ is on, or is a descendant of any variables on, a causal path (excluding the descendants of T alone)
  - the variables in $\boldsymbol{W}$ block *(in the sense of graphical models)* all the non-causal paths between T and Y
  - the variables in $\boldsymbol{W}$ block *(in the sense of graphical models)* all the non-causal paths between T and M
  - the variables in $\boldsymbol{W}$ block *(in the sense of graphical models)* all the non-causal paths between M and Y