Causal and Marginal Models

Robin Evans, University of Oxford
Vanessa Didelez, University of Bristol

University of Bristol
29th April 2016

Outline

Marginal Models

There are many situations in which we need to model marginal structure as part of a larger multivariate model:

- to account for dependence between individuals in panel studies;
- to enforce stationarity in longitudinal models;
- to model a marginal or conditional independence in a Bayesian network;
- in causal models;
- to transfer information across studies;
- etc...

Political Orientation

This is data from a panel study which rated political orientation from 1 (extremely liberal) to 7 (extremely conservative) in 1992 and 1994. (Bergsma, Croon and Hagenaars, 2013)

		Political Orientation 1994							
		1	2	3	4	5	6	7	**Total**
1992	1	3	4	1	2	0	1	0	11
	2	2	23	15	6	0	2	0	48
	3	1	8	23	9	9	1	0	51
	4	0	6	17	56	19	13	2	113
	5	0	1	1	18	40	29	3	92
	6	0	1	1	4	13	51	7	77
	7	0	0	0	0	2	11	3	16
	Total	6	43	58	95	83	107	16	408

Political Orientation

This is data from a panel study which rated political orientation from 1 (extremely liberal) to 7 (extremely conservative) in 1992 and 1994. (Bergsma, Croon and Hagenaars, 2013)

		Political Orientation 1994							
		1	2	3	4	5	6	7	**Total**
1992	1	3	4	1	2	0	1	0	11
	2	2	23	15	6	0	2	0	48
	3	1	8	23	9	9	1	0	51
	4	0	6	17	56	19	13	2	113
	5	0	1	1	18	40	29	3	92
	6	0	1	1	4	13	51	7	77
	7	0	0	0	0	2	11	3	16
	Total	6	43	58	95	83	107	16	408

Naïve modelling gives no evidence to reject null hypothesis of no change; full modelling gives strong evidence of a shift.

Parameterisations

A simple parameterisation just uses the cell probabilities $p_{11}, p_{21}, \ldots$.

		Y			
		1	2	3	**Total**
X	1	p_{11}	p_{12}	p_{13}	p_{1+}
	2	p_{21}	p_{22}	p_{23}	p_{2+}
	3	p_{31}	p_{32}	p_{33}	p_{3+}
	Total	p_{+1}	p_{+2}	p_{+3}	p_{++}

To model the marginals, we instead want to start with the row and column totals: $p_{1+}, p_{2+}, \ldots, p_{+1}, p_{+2} \ldots$.

On top of this, we need. . .

Parameterisations

A simple parameterisation just uses the cell probabilities $p_{11}, p_{21}, \ldots$.

		Y			
		1	2	3	**Total**
X	1	p_{11}	p_{12}	p_{13}	p_{1+}
	2	p_{21}	p_{22}	p_{23}	p_{2+}
	3	p_{31}	p_{32}	p_{33}	p_{3+}
	Total	p_{+1}	p_{+2}	p_{+3}	p_{++}

To model the marginals, we instead want to start with the row and column totals: $p_{1+}, p_{2+}, \ldots, p_{+1}, p_{+2} \ldots$.

On top of this, we need... the **odds ratios**!

$$\text{e.g.} \qquad \frac{p_{11}p_{22}}{p_{12}p_{21}}.$$

Smooth, Variation Independent Parameters

		Y			
		1	2	3	**Total**
X	1	p_{11}	p_{12}	p_{13}	p_{1+}
	2	p_{21}	p_{22}	p_{23}	p_{2+}
	3	p_{31}	p_{32}	p_{33}	p_{3+}
	Total	p_{+1}	p_{+2}	p_{+3}	p_{++}

- p_{1+}, p_{2+};
- p_{+1}, p_{+2};
- $$\frac{p_{11}p_{22}}{p_{12}p_{21}}, \; \frac{p_{11}p_{23}}{p_{13}p_{21}} \; \frac{p_{11}p_{32}}{p_{12}p_{31}}, \; \frac{p_{11}p_{33}}{p_{13}p_{31}}.$$

Smooth, Variation Independent Parameters

		Y			
		1	2	3	**Total**
X	1	p_{11}	p_{12}	p_{13}	p_{1+}
	2	p_{21}	p_{22}	p_{23}	p_{2+}
	3	p_{31}	p_{32}	p_{33}	p_{3+}
	Total	p_{+1}	p_{+2}	p_{+3}	p_{++}

- p_{1+}, p_{2+};
- p_{+1}, p_{+2};
- $$\frac{p_{11}p_{22}}{p_{12}p_{21}}, \ \frac{p_{11}p_{23}}{p_{13}p_{21}} \ \frac{p_{11}p_{32}}{p_{12}p_{31}}, \ \frac{p_{11}p_{33}}{p_{13}p_{31}}.$$

Using marginal probabilities and odds ratios is a

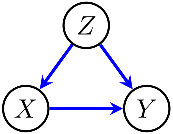
- **smooth** and
- **variation independent**

parameterisation.

Margins

Causal modelling typically involves marginal parameters with some more complicated global structure.

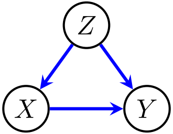
Suppose we want to know how X affects Y, but we have a measured confounder Z.



Margins

Causal modelling typically involves marginal parameters with some more complicated global structure.

Suppose we want to know how X affects Y, but we have a measured confounder Z.

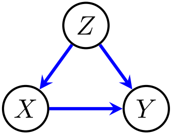


$$P(y \mid do(x)) = \sum_z P(z) P(y \mid x, z).$$

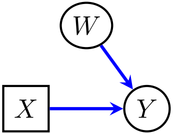
This quantity tells us *what would* happen if X were chosen independently of Z. How do we use it to generate observations from the real distribution with this marginal piece?

Multiple Experiments and Transportability

Suppose we have two experiments on some of the same variables:



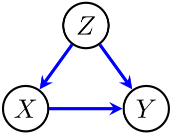
observational study, P



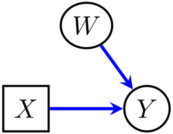
randomised trial, Q

Multiple Experiments and Transportability

Suppose we have two experiments on some of the same variables:



observational study, P



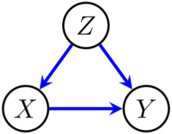
randomised trial, Q

Suppose want to assert / test that

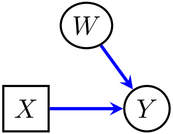
$$P(Y \mid do(X)) = Q(Y \mid do(X)).$$

Multiple Experiments and Transportability

Suppose we have two experiments on some of the same variables:



observational study, P



randomised trial, Q

Suppose want to assert / test that

$$P(Y \mid do(X)) = Q(Y \mid do(X)).$$

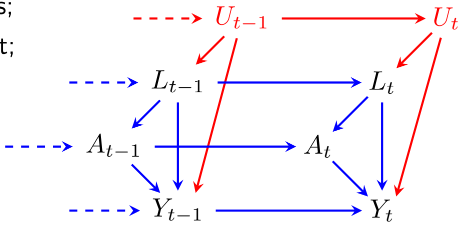
These are marginal parameters:

$$P(Y \mid do(X)) = \sum_Z P(Y \mid X, Z) \cdot P(Z)$$
$$Q(Y \mid do(X)) = Q(Y \mid X).$$

Survival Models

At each time t we have:

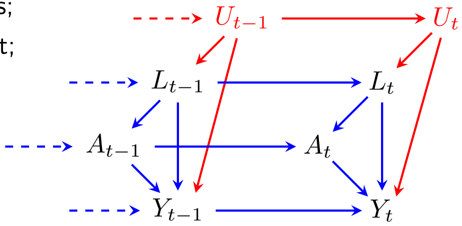
- U_t unobserved variables;
- L_t covariates;
- A_t treatment;
- Y_t survival.



Survival Models

At each time t we have:

- U_t unobserved variables;
- L_t covariates;
- A_t treatment;
- Y_t survival.



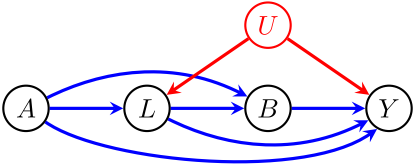
What is probability of survival ($Y = 1$) to next time point, given treatment?

$$P(Y_t = 1 \mid Y_{t-1} = 1, do(A_1, \ldots, A_t)).$$

Outline

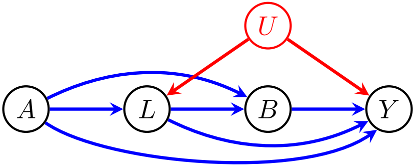
Causal Models

Take a simple two-step dynamic treatment model.



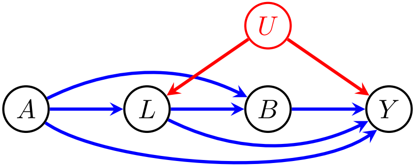
- A, B treatments (randomised);
- L intermediate outcome;
- Y final outcome;
- U unobserved confounders.

Identification



Question: how do the treatments causally affect the final outcome?
Or, if we treated everyone with (a, b), what would happen?

Identification

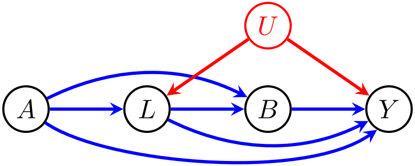


Question: how do the treatments causally affect the final outcome?
Or, if we treated everyone with (a, b), what would happen?

How do we identify this?

- $P(Y \mid A = a, B = b)$: ignoring/marginalising L;
- $P(Y \mid A = a, B = b, L = l)$: conditioning on L.

Identification



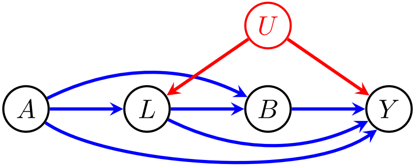
Question: how do the treatments causally affect the final outcome?
Or, if we treated everyone with (a, b), what would happen?

How do we identify this?

- $P(Y \mid A = a, B = b)$: ignoring/marginalising L;
- $P(Y \mid A = a, B = b, L = l)$: conditioning on L.

Neither has the desired causal interpretation!

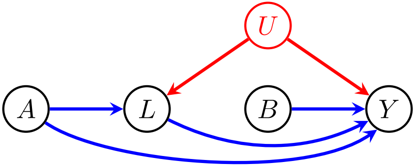
Identification



Can 'reweight' a sample/distribution to pretend that B was assigned independently of A and L:

$$P^*(A, L, B, Y) = P(A, L, B, Y)\frac{P(B)}{P(B \mid A, L)}.$$

Identification

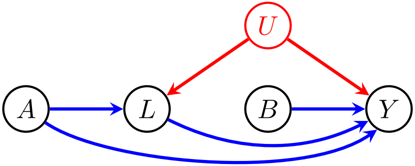


Can 'reweight' a sample/distribution to pretend that B was assigned independently of A and L:

$$P^*(A, L, B, Y) = P(A, L, B, Y)\frac{P(B)}{P(B \mid A, L)}.$$

In this new 'world', L is post-treatment, so just ignore it!
Then $P^*(Y \mid A, B)$ **does** have the desired causal interpretation!

Identification



Can 'reweight' a sample/distribution to pretend that B was assigned independently of A and L:

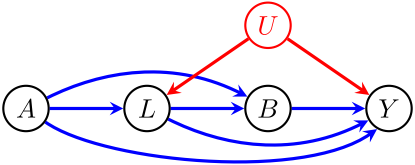
$$P^*(A, L, B, Y) = P(A, L, B, Y) \frac{P(B)}{P(B \mid A, L)}.$$

In this new 'world', L is post-treatment, so just ignore it!
Then $P^*(Y \mid A, B)$ **does** have the desired causal interpretation!

$$P^*(Y \mid A, B) = \sum_L P^*(L, Y \mid A, B) = \sum_L P(Y \mid A, L, B) \cdot P(L \mid A).$$

Parameterising Causal Models

Identification due to Robins (1986); more general results available (Shpitser and Pearl, 2006).



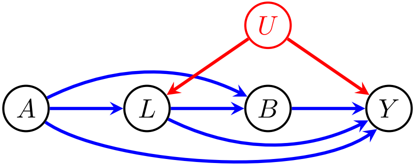
I will denote this causal quantity by

$$P(Y \mid do(A = a, B = b))$$

i.e. the distribution of outcome given that the treatments are *set* to (a, b) by intervention.

Parameterising Causal Models

Identification due to Robins (1986); more general results available (Shpitser and Pearl, 2006).



I will denote this causal quantity by

$$P(Y \mid do(A = a, B = b))$$

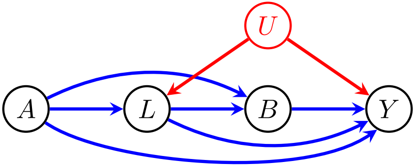
i.e. the distribution of outcome given that the treatments are *set* to (a, b) by intervention.

Causal question of interest might be:

"does $P(Y \mid do(A = a, B = b))$ depend upon a?"

Parameterising Causal Models

For likelihood-based inference and simulation, need a parametrisation.



Standard parameterisations lead to the **g-null paradox**.

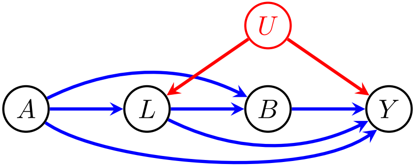
For example, if:

- logistic regression model for Y given A, B, L;
- linear Gaussian model for L given A;

then with faithfulness it is **impossible** for $P(Y \mid do(A, B))$ not to depend upon A. (Robins and Wasserman, 1997)

Parameterising Causal Models

For likelihood-based inference and simulation, need a parametrisation.



Standard parameterisations lead to the **g-null paradox**.

For example, if:

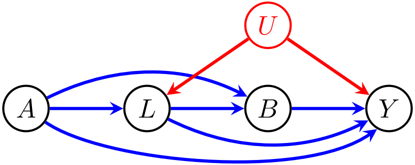
- logistic regression model for Y given A, B, L;
- linear Gaussian model for L given A;

then with faithfulness it is **impossible** for $P(Y \mid do(A, B))$ not to depend upon A. (Robins and Wasserman, 1997)

Naturally, this is disastrous for hypothesis testing.

Simulation

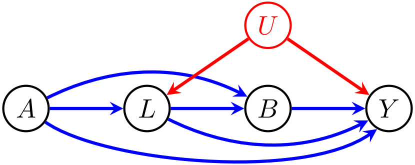
Havercroft and Didelez (2012) note that simulating data from this model such that $P(Y \mid do(A, B))$ independent of A is difficult in some cases.



Why?

Simulation

Havercroft and Didelez (2012) note that simulating data from this model such that $P(Y \mid do(A, B))$ independent of A is difficult in some cases.



Why?

Simulation requires $P(A, L, B, Y)$;
relationship to $P(Y \mid do(A, B))$ seems complicated.

g-null paradox shows we can't just specify a nice parametric model for P and then fix parameters until independence holds.

Recast the Problem

Define

$$P^*(Y, L \mid A, B) \equiv P(Y, L \mid do(A, B))$$
$$= P(Y \mid A, L, B) \cdot P(L \mid A).$$

Message: P^* is *just* a (conditional) probability distribution.

Recast the Problem

Define

$$P^*(Y, L \mid A, B) \equiv P(Y, L \mid do(A, B)) \\ = P(Y \mid A, L, B) \cdot P(L \mid A).$$

Message: P^* is *just* a (conditional) probability distribution.

Desired Properties of P^*

- nice model for $P^*(Y \mid A, L, B) = P(Y \mid A, L, B)$ for simulation.
- nice model for $P^*(Y \mid A, B)$ for statistical inference;
- nice model for $P^*(L \mid A, B) = P(L \mid A)$ to ensure $L \perp\!\!\!\perp B \mid A\, [P^*]$;

So how do we get this?

Recast the Problem

Define

$$P^*(Y, L \mid A, B) \equiv P(Y, L \mid do(A, B))$$
$$= P(Y \mid A, L, B) \cdot P(L \mid A).$$

Message: P^* is *just* a (conditional) probability distribution.

Desired Properties of P^*

- nice model for $P^*(Y \mid A, L, B) = P(Y \mid A, L, B)$ for simulation.
- nice model for $P^*(Y \mid A, B)$ for statistical inference;
- nice model for $P^*(L \mid A, B) = P(L \mid A)$ to ensure $L \perp\!\!\!\perp B \mid A\,[P^*]$;

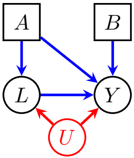
So how do we get this?

Short answer: **we can't**! It doesn't make sense to try to specify $P^*(Y \mid A, L, B)$ and $P^*(Y \mid A, B)$ separately.

Margins

A better way to think about this: given intervention distribution P^* suppose we have:

- a model for $P^*(Y \mid A, B)$;
- a model for $P^*(L \mid A, B) = P(L \mid A)$;

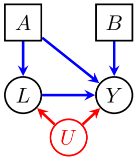


These do not fully specify $P^*(Y, L \mid A, B)$
so what else do we need?

Margins

A better way to think about this: given intervention distribution P^* suppose we have:

- a model for $P^*(Y \mid A, B)$;
- a model for $P^*(L \mid A, B) = P(L \mid A)$;



These do not fully specify $P^*(Y, L \mid A, B)$ so what else do we need?

Answer

The Y-L odds ratio, conditional on $A = a, B = b$:

$$\phi_{YL}(Y, L \mid A, B).$$

The additional information given by $P(Y \mid A, L, B)$ is then **redundant**.

Outline

Generalising Odds Ratios

Familiar definition of an odds ratio:

$$OR(X,Y) = \frac{P(X=1,Y=1) \cdot P(X=0,Y=0)}{P(X=1,Y=0) \cdot P(X=0,Y=1)}.$$

Generalising Odds Ratios

Familiar definition of an odds ratio:

$$OR(X,Y) = \frac{P(X=1,Y=1)\cdot P(X=0,Y=0)}{P(X=1,Y=0)\cdot P(X=0,Y=1)}.$$

Information contained is the same as:

$$\phi_{XY}(x,y) \equiv P(X=x,Y=y)\cdot u(x)\cdot v(y).$$

for unknown functions $u, v > 0$.

Generalising Odds Ratios

Familiar definition of an odds ratio:

$$OR(X,Y) = \frac{P(X=1,Y=1) \cdot P(X=0,Y=0)}{P(X=1,Y=0) \cdot P(X=0,Y=1)}.$$

Information contained is the same as:

$$\phi_{XY}(x,y) \equiv P(X=x,Y=y) \cdot u(x) \cdot v(y).$$

for unknown functions $u, v > 0$.

Can obtain the familiar odds ratio by taking the cross-ratio:

$$\frac{\phi_{XY}(1,1) \cdot \phi_{XY}(0,0)}{\phi_{XY}(1,0) \cdot \phi_{XY}(0,1)} = \frac{P(X=1,Y=1) \cdot P(X=0,Y=0)}{P(X=1,Y=0) \cdot P(X=0,Y=1)}.$$

Generalising Odds Ratios

Let p be a density for X, Y.

The **odds ratio** for X, Y is the equivalence class of functions ϕ_{XY} such that

$$\phi_{XY}(x, y) = p(x, y) \cdot u(x) \cdot v(y).$$

some functions $u, v > 0$.

Generalising Odds Ratios

Let p be a density for X, Y.

The **odds ratio** for X, Y is the equivalence class of functions ϕ_{XY} such that

$$\phi_{XY}(x, y) = p(x, y) \cdot u(x) \cdot v(y).$$

some functions $u, v > 0$.

Some points:

- defined for any distribution with a density;
- p is a member of the equivalence class;
- there's no requirement for p to be positive;
- iterative proportional fitting recovers the joint distribution.

Specifying Margins

Let $r_{XY}(x,y)$ be a joint distribution with odds ratio ϕ_{XY}.

Theorem

Let p_X and p_Y be densities such that $p_X \ll r_X$ and $p_Y \ll r_Y$. Then there exists a unique joint distribution with margins p_X, p_Y and odds ratio ϕ_{XY}.

This follows from Csiszár (1975).

Specifying Margins

Let $r_{XY}(x, y)$ be a joint distribution with odds ratio ϕ_{XY}.

Theorem

Let p_X and p_Y be densities such that $p_X \ll r_X$ and $p_Y \ll r_Y$. Then there exists a unique joint distribution with margins p_X, p_Y and odds ratio ϕ_{XY}.

This follows from Csiszár (1975).

This is a form of **variation independence**: we can paste together essentially any dependence structure with any margins and get a distribution.

Examples

- For discrete variables this reduces to the 'usual' odds ratio;
- for Gaussian variables:

$$\phi_{XY} \sim \exp\left(\frac{\rho x y}{\sigma_x \sigma_y (1-\rho^2)}\right)$$

- multivariate t-distribution ($\boldsymbol{x} = (x, y)^T$):

$$\phi_{XY} \sim \left(1 + \nu^{-1} \boldsymbol{x}^T \Sigma^{-1} \boldsymbol{x}\right)^{-\nu/2-1}$$

Copulae

A popular way of modelling dependence between variables without specifying margins is to use a **copula** model.

$$P(X \leq x, Y \leq y) = C(F_X(x), F_Y(y))$$

$$p(x, y) = c(F_X(x), F_Y(y)) \cdot p(x) \cdot p(y).$$

Note we can't obtain the copula from the odds ratio (or vice versa) without knowing the margins.

Copulae

A popular way of modelling dependence between variables without specifying margins is to use a **copula** model.

$$P(X \leq x, Y \leq y) = C(F_X(x), F_Y(y))$$

$$p(x, y) = c(F_X(x), F_Y(y)) \cdot p(x) \cdot p(y).$$

Note we can't obtain the copula from the odds ratio (or vice versa) without knowing the margins.

Features:

- Well established methods for fitting, simulation, inference.
- Non-continuous variables lead to unidentifiability.
- Not clear which is easier to interpret, copula or odds ratio.

Outline

Cognate Probabilities

We say a quantity $f(x_C \mid x_A)$ is a **cognate distribution** to the conditional probability density $p(x_C \mid x_A)$ if is of the form

$$f(x_C \mid x_A) \equiv \int_{x_B} p(x_C \mid x_A, x_B) \cdot w(x_B \mid x_A) \, dx_B,$$

where $w(x_B \mid x_A) > 0$ is a function of $p(x_A, x_B)$ such that $\int w(x_B \mid x_A) \, dx_B = 1$ for each x_A.

Cognate Probabilities

We say a quantity $f(x_C \mid x_A)$ is a **cognate distribution** to the conditional probability density $p(x_C \mid x_A)$ if is of the form

$$f(x_C \mid x_A) \equiv \int_{x_B} p(x_C \mid x_A, x_B) \cdot w(x_B \mid x_A) \, dx_B,$$

where $w(x_B \mid x_A) > 0$ is a function of $p(x_A, x_B)$ such that $\int w(x_B \mid x_A) \, dx_B = 1$ for each x_A.

Examples

$$\begin{aligned} p(x_C \mid x_A) &= \int p(x_C \mid x_A, x_B) \cdot p(x_B \mid x_A) \, dx_B \\ p(x_C \mid do(x_A)) &= \int p(x_C \mid x_A, x_B) \cdot p(x_B) \, dx_B \\ \mathbb{E}X_C(x_A, x'_A) &= \int p(x_C \mid x_A, x_B) \cdot p(x_B \mid x'_A) \, dx_B. \end{aligned}$$

Cognate Probabilities

In the discrete case, we can substitute cognate quantities in parameterisations without any problem.

Theorem

Suppose we have a multivariate discrete parameterisation which is hierarchical and consists of probabilities of the form $P(X_{A_i} \mid X_{B_i})$.

Then if we replace any of these $P^*(X_{A_i} \mid X_{B_i})$ with cognate quantities, the parameterisation is still smooth.

Conversely, a parameterisation involving two quantities which *are* cognate to one another will not be smooth, and any parameterisation which contrasts cognate quantities is non-smooth.

Cognate Probabilities

In the discrete case, we can substitute cognate quantities in parameterisations without any problem.

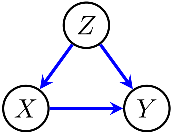
Theorem

Suppose we have a multivariate discrete parameterisation which is hierarchical and consists of probabilities of the form $P(X_{A_i} \mid X_{B_i})$.

Then if we replace any of these $P^*(X_{A_i} \mid X_{B_i})$ with cognate quantities, the parameterisation is still smooth.

Conversely, a parameterisation involving two quantities which *are* cognate to one another will not be smooth, and any parameterisation which contrasts cognate quantities is non-smooth.

Cognate Probabilities

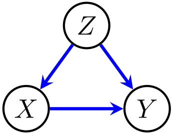


Example.

Since

$P(Z)$ $\quad P(X \mid Z)$ $\quad P(Y \mid X)$ $\qquad \phi_{YZ|X}$ is smooth...

Cognate Probabilities



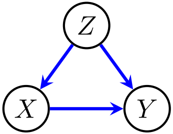
Example.

Since

$$P(Z) \qquad P(X \mid Z) \qquad P(Y \mid X) \qquad \phi_{YZ|X}$$ is smooth...

$$P(Z) \qquad P(X \mid Z) \qquad P(Y \mid do(X)) \qquad \phi_{YZ|X}$$ is also smooth!

Cognate Probabilities



Example.

Since

$P(Z)$	$P(X \mid Z)$	$P(Y \mid X)$	$\phi_{YZ\mid X}$	is smooth...
$P(Z)$	$P(X \mid Z)$	$P(Y \mid do(X))$	$\phi_{YZ\mid X}$	is also smooth!

Example.

$$\theta_x = P(Y = 1 \mid do(X = x)) - P(Y = 1 \mid X = x)$$

cannot form part of a smooth parameterisation.

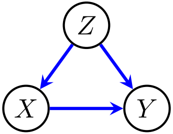
Results

Theorem

Consider (possibly vector valued) Y and X, Z. Then can parameterise joint distribution $P(Z, X, Y)$ with:

$$P(Z, X) \qquad \underbrace{P^*(Y \mid X)}_{\text{cognate to } P} \qquad \underbrace{\phi_{ZY}(Z, Y \mid X)}_{P-\text{odds ratio}}$$

and these three pieces are variation independent of one another.



Results

It follows that, given a temporal ordering, we can apply this result inductively, obtaining one ‘piece of information of interest’ for each variable.

Suppose we have an ordered set of variables: $X_1, X_2, \ldots, X_k$.

For each X_i divide the predecessors into two sets: $\boldsymbol{W}_i \cup \boldsymbol{V}_i$.

Theorem

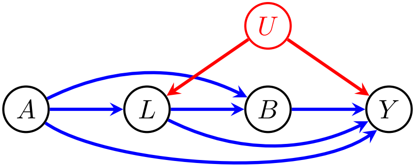
We can obtain a variation independent parameterisation which includes

$$P^*(X_i \mid \boldsymbol{W}_i) \qquad \forall i.$$

for any set of cognate quantities $P^*(X_i \mid \boldsymbol{W}_i)$.

Outline

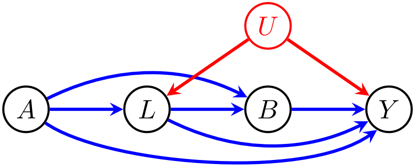
A Recipe



For our problem, separately specify (nice, parametric) models for:

- $P(A)$;
- $P(L \mid A)$;
- $P(B \mid A, L)$;

A Recipe



For our problem, separately specify (nice, parametric) models for:

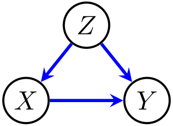
- $P(A)$;
- $P(L \mid A)$;
- $P(B \mid A, L)$;
- $P(Y \mid do(A, B))$ and $\phi_{YL}(Y, L \mid A, B)$ (the conditional odds ratio).

This is a fully variation independent, with no redundancy.

Example: Observed Confounding

Can parameterize this as:

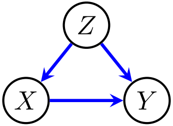
- $P(Z)$, $P(X \mid Z)$;
- $P(Y \mid do(X))$, $\phi_{YZ}(Y, Z \mid X)$.



Example: Observed Confounding

Can parameterize this as:

- $P(Z)$, $P(X \mid Z)$;
- $P(Y \mid do(X))$, $\phi_{YZ}(Y, Z \mid X)$.



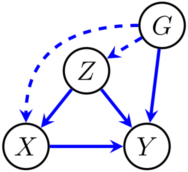
The variation independence is useful:

- easy to incorporate covariates in GLM form;
- no danger of choosing impossible higher order interactions (so no g-null paradox!);
- means independent priors are valid.

Example: Observed Confounding

Can parameterize this as:

- $P(Z)$, $P(X \mid Z)$;
- $P(Y \mid do(X))$, $\phi_{YZ}(Y, Z \mid X)$.



The variation independence is useful:

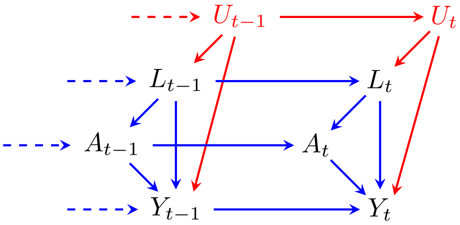
- easy to incorporate covariates in GLM form;
- no danger of choosing impossible higher order interactions (so no g-null paradox!);
- means independent priors are valid.

Example, suppose want to model how is causal effect of X on Y modulated by G. Then we can do this with a logistic regression form:

$$\operatorname{logit} P(Y = 1 \mid do(X), G) = f(X, G).$$

Example: Survival Models

Young and Tchetgen Tchetgen (2014) consider survival models:



What is probability of survival ($Y = 1$) to next time point, given treatment?

$$P(Y_t = 1 \mid Y_{t-1} = 1, do(A_1, \dots, A_t)).$$

No problem! What remains is the dependence structure between L's and Y_t given $A_1, \dots, A_t$

Example: Survival Models

Hence simulation in some cases becomes relatively easy under a null; e.g.:

$$P(Y_t \mid Y_{t-1} = 1, do(A_1, \ldots, A_t)) = P(Y_t \mid Y_{t-1} = 1).$$

Young and Tchetgen Tchetgen note that this is not at all trivial.

Example: Survival Models

Hence simulation in some cases becomes relatively easy under a null; e.g.:

$$P(Y_t \mid Y_{t-1} = 1, do(A_1, \ldots, A_t)) = P(Y_t \mid Y_{t-1} = 1).$$

Young and Tchetgen Tchetgen note that this is not at all trivial.

Can also easily incorporate, for e.g., a **stationarity assumption**.

$$P(Y_t \mid Y_{t-1} = 1, do(A_t = a)) = g(a)$$

What Can't Be Done

With each parameter (either conditional distribution or odds ratio) we can associate a collection of subsets of variables:

$$\mathbb{D}(P(X_A \mid X_B)) = \{W \subseteq A \cup B : A \cap W \neq \emptyset\}$$

$$\mathbb{D}(\phi_{AB}(X_A, X_B \mid X_C)) = \{W \subseteq A \cup B \cup C : A \cap W \neq \emptyset, B \cap W \neq \emptyset\}.$$

What Can't Be Done

With each parameter (either conditional distribution or odds ratio) we can associate a collection of subsets of variables:

$$\mathbb{D}(P(X_A \mid X_B)) = \{W \subseteq A \cup B : A \cap W \neq \emptyset\}$$

$$\mathbb{D}(\phi_{AB}(X_A, X_B \mid X_C)) = \{W \subseteq A \cup B \cup C : A \cap W \neq \emptyset, B \cap W \neq \emptyset\}.$$

Examples:

$$\mathbb{D}(P(Y \mid X, Z)) = \{\{Y\},\ \{X, Y\},\ \{Y, Z\},\ \{X, Y, Z\}\}$$

$$\mathbb{D}(\phi_{YZ}(Y, Z \mid X)) = \{\{Y, Z\},\ \{X, Y, Z\}\}.$$

What Can't Be Done

Proposition

Let ψ, ψ' be two parameters (i.e. cognate conditional distributions or odds ratios).
If $\mathbb{D}(\psi) \cap \mathbb{D}(\psi') \neq \emptyset$ then any parameterisation that includes ψ and ψ' is non-smooth.

What Can't Be Done

Proposition

Let ψ, ψ' be two parameters (i.e. cognate conditional distributions or odds ratios).
If $\mathbb{D}(\psi) \cap \mathbb{D}(\psi') \neq \emptyset$ then any parameterisation that includes ψ and ψ' is non-smooth.

Example

In our original example we tried to have:

$$P(Y \mid A, B, L) \qquad P(Y \mid do(A, B)).$$

But these both include $\{Y\}$, $\{Y, A\}$, $\{Y, B\}$, $\{Y, A, B\}$.

Example: History-Adjusted Marginal Structural Models

Denote $\overline{A}_t = (A_1, \dots, A_t)$ and $\underline{A}_t = (A_t, \dots, A_T)$.

van der Laan at al (2005) introduce a **history-adjusted** model that models:

$$p(Y \mid \overline{L}_t, \overline{A}_{t-1}, do(\underline{A}_t)), \qquad t = 1, \dots, T.$$

Example: History-Adjusted Marginal Structural Models

Denote $\overline{A}_t = (A_1, \dots, A_t)$ and $\underline{A}_t = (A_t, \dots, A_T)$.

van der Laan at al (2005) introduce a **history-adjusted** model that models:

$$p(Y \mid \overline{L}_t, \overline{A}_{t-1}, do(\underline{A}_t)), \qquad t = 1, \dots, T.$$

By the previous result, we cannot expect to model e.g.

$$p(Y \mid L_1, do(A_1, A_2)) \qquad p(Y \mid L_1, L_2, A_1, do(A_2)).$$

separately.

The incompatibility of the models used was pointed out by Robins, Hernán and Rotnitzky (2007).

Summary

- **Causal models are marginal models**;

Summary

- **Causal models are marginal models**;
- the g-null paradox arises from trying to specify the same quantity twice;
- this can be avoided by understanding which parameters are ‘free’ to be specified;

Summary

- **Causal models are marginal models**;
- the g-null paradox arises from trying to specify the same quantity twice;
- this can be avoided by understanding which parameters are ‘free’ to be specified;
- application to marginal structural models, survival models, stationarity, transportability...

Summary

- **Causal models are marginal models**;
- the g-null paradox arises from trying to specify the same quantity twice;
- this can be avoided by understanding which parameters are ‘free’ to be specified;
- application to marginal structural models, survival models, stationarity, transportability...
- simulation becomes much easier in Gaussian, discrete cases, some copula models;
- there is a large literature on marginal models to look at for other cases.

Thank you!

References

Bergsma and Rudas. Marginal log-linear parameters, *Ann. Statist.*, 2002.

Bergsma, Croon, Hagenaars. Advancements in Marginal Modeling for Categorical Data, *Sociological Methodology*, 2013.

Csiszár. I-Divergence Geometry of Probability Distributions and Minimization Problems, *Ann. Prob.*, 1975.

Evans and Richardson. Marginal log-linear parameters for acyclic directed mixed graphs, *JRSS-B*, 2013.

Havercroft and Didelez. Simulating from marginal structural models with time-dependent confounding, *Stat. Med.*, 2012.

Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect, *Math. Modelling*, 1986.

Robins, Hernán, Rotnitzky. Invited Commentary: Effect Modification by Time-varying Covariates, *Am. J. Epi.*, 2007.

Robins and Wasserman. Estimation of Effects of Sequential Treatments by Reparameterizing DAGs, *UAI*, 1997.

Shpitser and Pearl, Identification of Joint Interventional Distributions in Recursive Semi-Markovian Causal Models, *AAAI*, 2006.

van der Laan, Petersen and Joffe. History-Adjusted Marginal Structural Models and Statically-Optimal Dynamic Treatment Regimens, *Biostatistics*, 2005.

Young and Tchetgen Tchetgen. Simulation from a known Cox MSM using standard parametric models for the g-formula, *Stat. Med.*, 2014.