Parameterizing and Simulating from Causal Models

Robin J. Evans
University of Oxford

Vanessa Didelez
Leibniz Institute BIPS, and University of Bremen

RSS Discussion Meeting
3rd October 2023

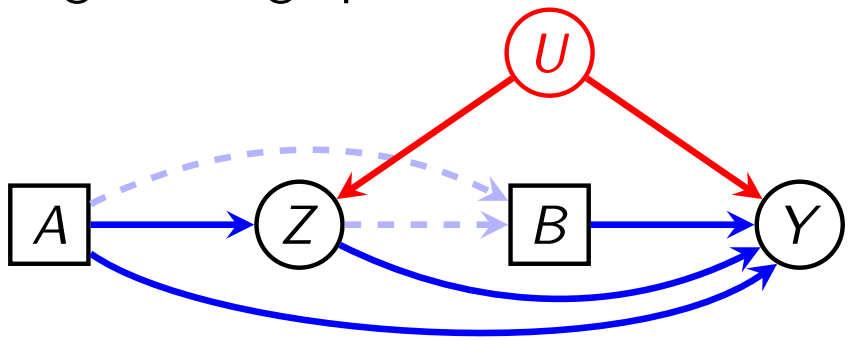
Outline

1. Motivation
2. Marginal Models
3. Frugal Parameterization and Main Results
4. Simulations
5. Other Applications

Outline

A Dynamic Treatment Model

Consider the following causal graph.



Here:

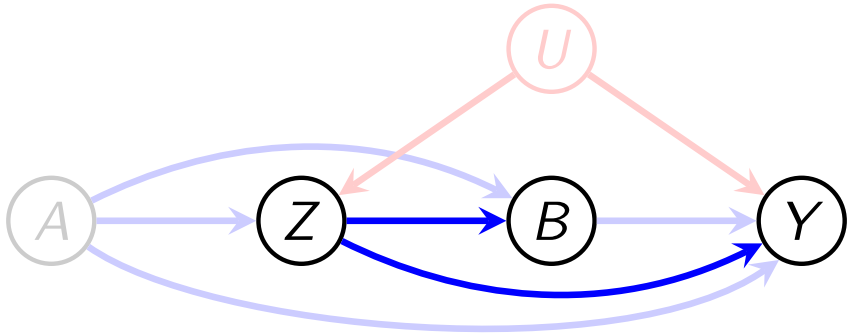
- A, B are treatments given by a doctor;
- Z is an intermediate outcome;
- Y is a final outcome;
- U represents unobserved confounders.

Suppose we're interested in interventions on A and B.

What would happen if **everyone** were given treatments $(A, B) = (a, b)$?
i.e. we want

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

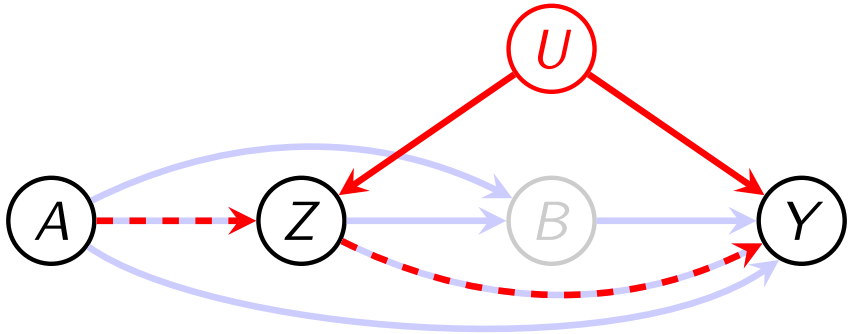
Identification



How can we identify $P(Y \mid do(A = a, B = b))$?

Just regressing on the treatments fails, because Z is a confounder of the causal effect of B on Y.

Identification



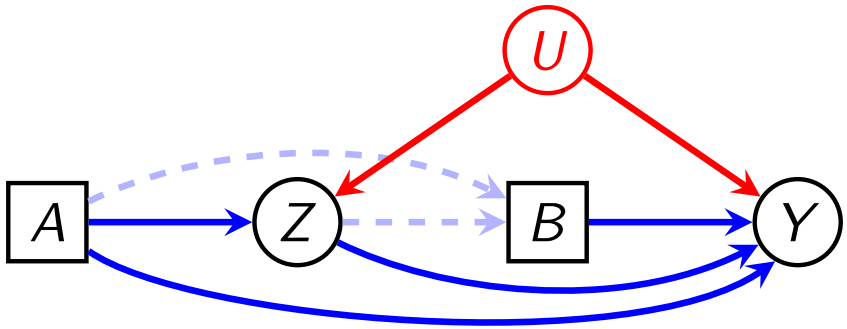
How can we identify $P(Y \mid do(A = a, B = b))$?

Just regressing on the treatments fails, because Z is a confounder of the causal effect of B on Y.

Regressing on the treatments and covariates **also** fails, as Z is a mediator of the effect of A on Y and a collider opening a non-causal path.

Here, Z is a **time-varying confounder** (aka 'treatment-confounder feed-back').

Identification



Under 'standard' causal assumptions, we can identify $P(Z, Y \mid do(A = a, B = b))$ using **inverse probability weighting**:

$$p(z, y \mid do(a, b)) = \frac{p(a, z, b, y)}{p(a) \cdot p(b \mid a, z)}.$$

Alternatively, we can use the **g-formula** (Robins, 1986):

$$p(z, y \mid do(a, b)) = p(z \mid a) \cdot p(y \mid a, z, b).$$

Marginalizing over z then yields:

$$p(y \mid do(a, b)) = \int_{\mathcal{Z}} \frac{p(a, z, b, y)}{p(a) \cdot p(b \mid a, z)} \, dz = \int_{\mathcal{Z}} p(z \mid a) \cdot p(y \mid a, z, b) \, dz.$$

Marginal Structural Models (MSMs)

Models of $p(y|do(a,b))$ are **marginal structural models** (Robins, 2000).

MSMs are very popular in epidemiology, as **time-varying confounding** is simply 'removed' by suitable weighting.

Examples:

1. ART therapy ('when to start', 'when to switch') for HIV-patients with CD4 count as time-varying confounder;
2. survival of Cystic Fibrosis patients under sustained treatments ('always' vs. 'never');
3. cancer-screening attendance ('regular' vs. 'delayed' vs. 'never') with cancer incidence or mortality outcomes;
4. assessing side effects of (sustained or combined) anti-diabetic drug use in type-2 diabetes patients.

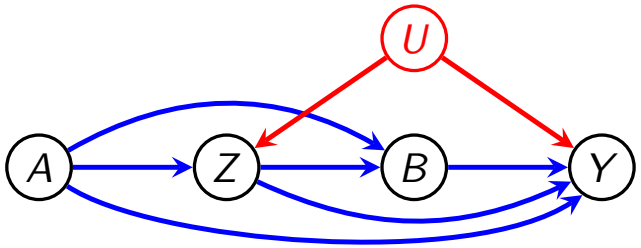
Marginal Structural Models (MSMs)

However, papers that **simulate** from them do so in **indirect** ways.

Examples

- Young et al. (2008, 2010) give two different approaches to simulating from MSMs using special cases of other models and computing the implied parameters.
- Havercroft and Didelez (2012) try simulating such that $p(y \mid do(a, b))$ does not depend upon a. Their approach requires removing the direct effect $Z \to Y$.
- Keogh et al. (2021) use fully conditional (additive or Cox) hazard models and then work out the implied MSMs.

Obstacles



Simulating from MSMs is hard.

Why?

In discussing marginal structural models Robins (2000, p107) notes:

> *"...the difficulty in performing likelihood-based inference... since the likelihood is a* **computational nightmare**.*"*

The g-formula (alternative to IPTW) suffers from the **g-null paradox** (Robins and Wasserman, 1997).

The G-null Paradox

Robins and Wasserman (1997) show that specifying seemingly nice parametric models for $Z \mid A$ and $Y \mid A, Z, B$ lead to it being (almost) **impossible** for the null hypothesis to hold.

Suppose
$$Z \mid A \sim \text{Bernoulli}(\text{expit}(\alpha A))$$
$$\mathbb{E}[Y \mid A, Z, B] = A\beta_a + Z\beta_z + B\beta_b.$$

Then we can compute:

$$\mathbb{E}[Y \mid do(A, B)] = A\beta_a + B\beta_b + \text{expit}(\alpha A) \cdot \beta_z.$$

Hence g-null holds $\iff \beta_a = 0$ and $\alpha \cdot \beta_z = 0$

$$\iff \text{either } Y \perp\!\!\!\perp A, Z \mid B \text{ or } \left\{ \begin{array}{l} Z \perp\!\!\!\perp A \\ Y \perp\!\!\!\perp A \mid Z, B \end{array} \right\}.$$

This is **much more restrictive** than the hypothesis of interest.

Lesson: just specifying nice models for all conditionals is **not helpful** if we want data under a specific causal hypothesis.

Further Problems

Another approach to overcoming (some of) these problems is to use models that are **over-parameterized** and/or not **congenial**.

Example

Some propose to consider separate marginal and conditional specifications:

$$\text{logit}\,\mathbb{E}[Y \mid A] = \alpha_0 + \alpha_a A$$
$$\text{logit}\,\mathbb{E}[Y \mid A, B] = \beta_0 + \beta_a A + \beta_b B.$$

However, **only trivial models** satisfy these restrictions for any $\alpha_0, \alpha_a, \beta_0, \beta_a$ if $\beta_b \neq 0$, because the logit function is not **collapsible**.

Objectives

There are three broad objectives for our proposal.

1 Parameterization. Describe a joint distribution P of all variables (confounders, treatments, outcomes) that:

- obeys an **explicitly provided parameterization** of the causal functional of interest;
- allows the rest of the distribution to be as **flexible** as possible;
- remains **coherent** and **unambiguous**.

2 Simulation. Obtain samples from P such that:

- observationally the data exhibit **complex confounding** structures;
- under an intervention they **obey the causal model**.

3 Fitting. Allow for fitting typically performed semi-paramerically (e.g. marginal structural models), using **likelihood-based methods**.

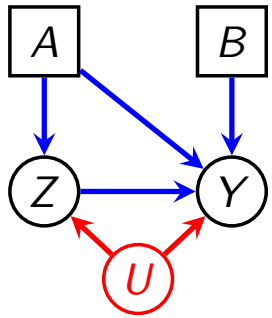
Outline

Marginal Models

Define $$p^*(z, y \mid a, b) \equiv p(z, y \mid do(a, b)) = p(z \mid a) \cdot p(y \mid a, z, b).$$

Given interventional distribution P^* suppose we have:

- a model for $p^*(y \mid a, b)$;
- a model for $p^*(z \mid a, b) = p(z \mid a)$.



Question

These do not fully specify $p^*(z, y \mid a, b)$ so what else do we need?

Answer

Some sort of (conditional) **dependence measure** for Y and Z under P^* (e.g. a conditional odds ratio or copula):

$$\phi^*_{ZY|AB}(z, y \mid a, b).$$

Marginal Tension

We've seen that there is generally a tension between:

- simple specification of the **joint distribution** P, in order to facilitate simulation and likelihood-based inference;
- simple specification of the **target of inference** $p^*(y \mid a, b)$ (i.e. some interventional marginal quantity) in order that it is interpretable;
- enforcing marginal **constraints** implied by the causal model. (In our case this was $Z \perp\!\!\!\perp B \mid A$ under P^*.)

Our proposal resolves these as best one can.

Outline

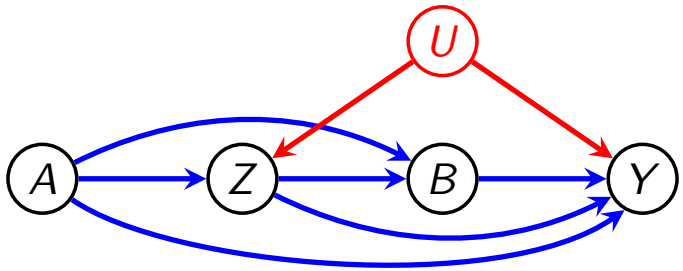
Setup

In general, we consider three groups of variables:

X	treatments and effect modifiers
Y	outcome(s) of interest
Z	other variables to be marginalized

These can all be **vector valued**.

Note that there is not necessarily a strict causal order on Z, X and Y: in our example, we had $X = (A, B)$.



Object of interest is $p^*(y \mid x)$ for some interventional P^*.

Cognate Distributions

We need a **marginal** parameterization, but some of the 'margins' we are interested in are non-standard.

Let $w(z \mid x)$ be a **kernel function**:

- $w(z \mid x) \geq 0$;
- $\int_{\mathcal{Z}} w(z \mid x)\, dz = 1$ for each x.

We allow it to be a (smooth) function of $p(z, x)$.

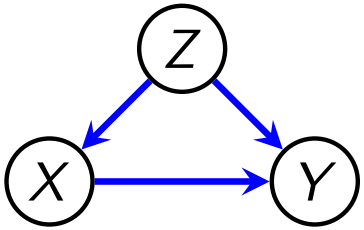
Definition

We say $p^*(y \mid x)$ is **cognate** to $p(y \mid x)$ (within $p(z, x, y)$) if

$$p^*(y \mid x) \equiv \int_{\mathcal{Z}} p(y \mid x, z) \cdot w(z \mid x)\, dz.$$

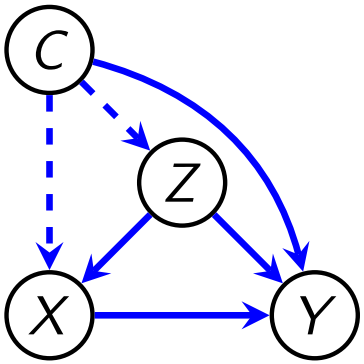
Cognate Distributions: Examples

Examples



$$p(y \mid x) = \int_{\mathcal{Z}} p(y \mid x, z) \cdot p(z \mid x)\, dz$$

$$p(y \mid do(x)) = \int_{\mathcal{Z}} p(y \mid x, z) \cdot p(z)\, dz$$



$$p(y \mid c; do(x)) = \int_{\mathcal{Z}} p(y \mid c, z, x) \cdot p(z \mid c)\, dz$$

$$p(Y(x) \mid x') = \int_{\mathcal{Z}} p(y \mid x, z) \cdot p(z \mid x')\, dz.$$

(Here $Y(x)$ is the **potential outcome** for Y when X is set to x.)

Note that $\mathbb{E}[Y(0) \mid X = 1]$ appears in the **effect of treatment on the treated** estimand.

Frugal Parameterization

Definition

A **frugal parameterization** consists of three separate (smooth and regular) parametric models for:

- $p(z, x)$ ('the past');
- $p^*(y \mid x)$ (distribution of interest);
- $\phi^*_{ZY|X}(z, y \mid x)$ (a dependence measure).

The distribution of interest can be any that is cognate to $p(y \mid x)$.

These quantities:

- specify the whole distribution P;
- can be chosen to be variation independent;
- have no redundancy.

Main Result

Theorem

Consider an outcome Y, and causally prior variables X and Z.
Then we can **smoothly parameterize** the joint distribution P with a frugal parameterization of

$$p(z, x) \qquad p^*(y \mid x) \qquad \phi^*_{ZY|X}(z, y \mid x),$$

where $p^*(y \mid x)$ is cognate to $p(y \mid x)$, if and only if P can also be smoothly parameterized by the same models applied to

$$p(z, x) \qquad p(y \mid x) \qquad \phi_{ZY|X}(z, y \mid x).$$

(That is, the ordinary conditional and the dependence measure in P.)

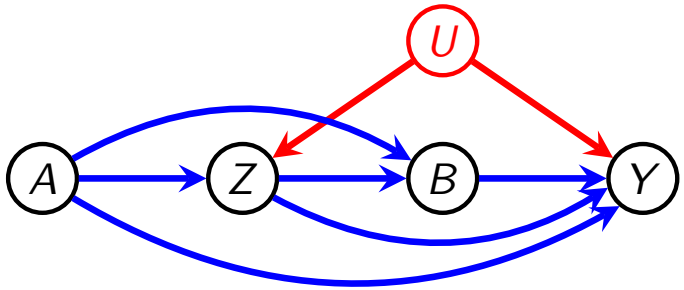
Can choose distinct parts of second list to be **variation independent**, in which case same is true for the first.

This gives us the **best of both worlds**: a coherent joint distribution and a marginal specification of our choice.

Outline

Copula Model Example

Consider the two-step dynamic model from Havercroft and Didelez (2012).



Then:

- Simulate $A, B \sim \text{Bernoulli}(\frac{1}{2})$ independently;
- Obtain conditional quantiles of Y, Z from a Gaussian copula with correlation $2\,\text{expit}(1 + a/2) - 1$;
- Using inversion, set $Z \mid A = a \sim \text{Exp}(\exp(0.2a - 0.3))$;
- Set $Y \mid do(A = a, B = b) \sim N(-0.5 + 0.2a + 0.3b,\ 1)$;

After rejection sampling:

- $B \mid A = a, Z = z \sim \text{Bernoulli}(\text{expit}(a/2 + z/2))$.

We take a sample of size $n = 10^4$ using the R package `causl` (Evans, 2021, `https://github.com/rje42/causl`).

Results (Outcome Regression)

We fit the naïve outcome regression model

$$\mathbb{E}[Y \mid A = a, B = b] = \beta_0 + \beta_A a + \beta_B b + \beta_{AB} ab.$$

	truth	estimate	bias	s.e.	z-value	p-value
intercept	-0.5	-0.564	-0.064	0.021	-3.12	1.7×10^{-3}
A	0.2	0.156	-0.044	0.030	-1.44	0.15
B	0.3	0.448	0.148	0.028	5.22	1.8×10^{-7}
$A \cdot B$	0.0	0.047	0.047	0.040	1.18	0.24

Results (IPW)

We estimate the inverse weights by fitting a logistic regression for $B \mid A, Z$:

$$\text{logit}\, P(B = 1 \mid A = a, Z = z) = \alpha_0 + \alpha_A a + \alpha_Z z + \alpha_{AZ} az.$$

We then **inverse weight** each observation by $\hat{p}(b \mid a, z)$, and fit the outcome model using these inverse weights.

The bias is very small:

	truth	estimate	bias	s.e.	z-value	p-value
intercept	-0.5	-0.489	0.011	0.021	0.49	0.62
A	0.2	0.196	-0.004	0.032	-0.14	0.89
B	0.3	0.302	0.002	0.029	0.08	0.94
$A \cdot B$	0.0	0.003	0.003	0.042	0.07	0.95

This suggests that (i) our simulation is working well; and (ii) IPW works.

Results (Maximum Likelihood Estimation)

Since we have a parametric model, we can also evaluate the likelihood and compute the MLE

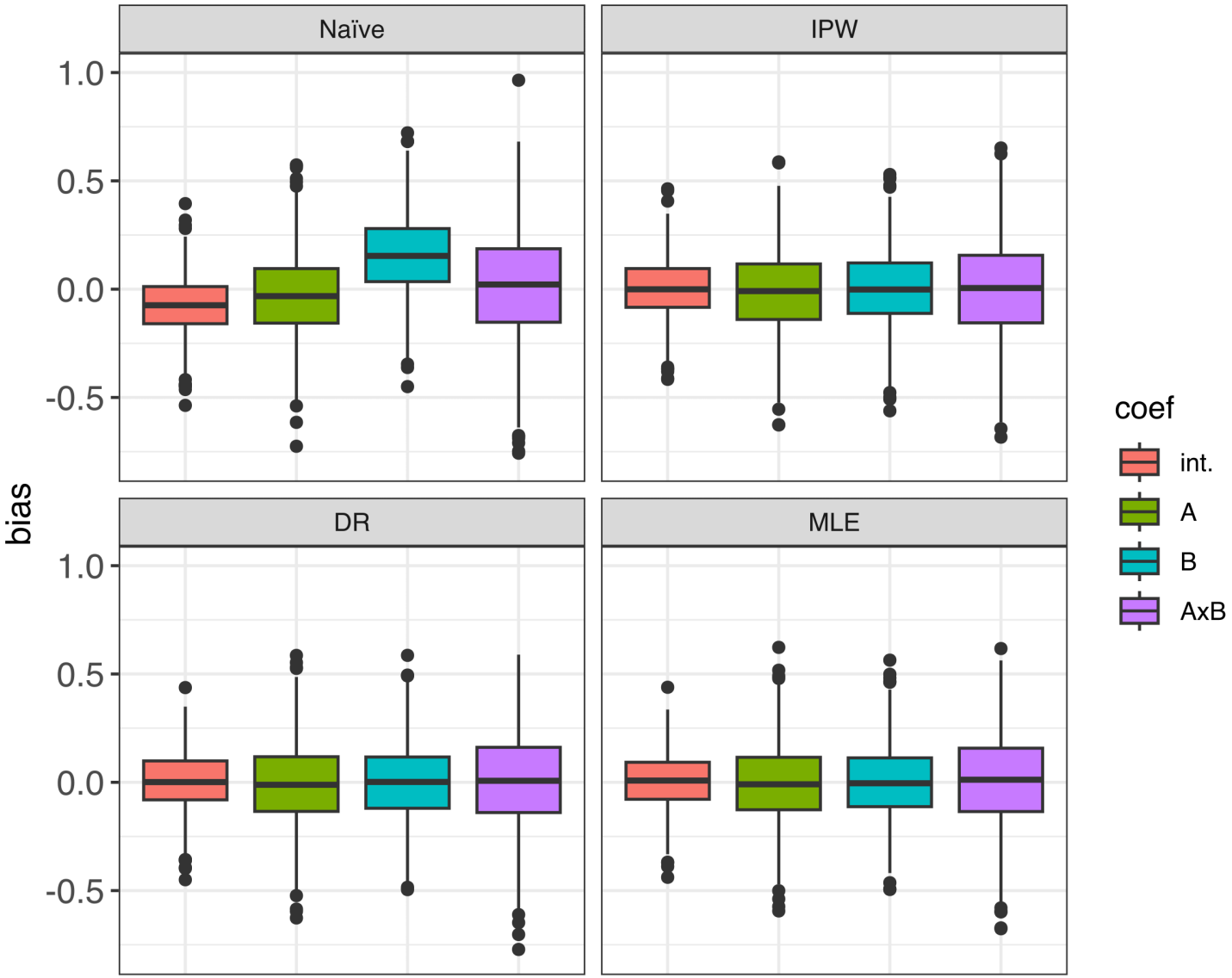
Comparison with MLE is very useful for simulation, because it provides an optimally **efficient** comparator.

	truth	estimate	bias	s.e.	z-value	p-value
intercept	-0.5	-0.490	0.010	0.019	0.53	0.60
A	0.2	0.195	-0.005	0.027	-0.20	0.84
B	0.3	0.302	0.002	0.026	0.08	0.94
$A \cdot B$	0.0	0.010	0.010	0.034	0.28	0.78

However, we don't recommended using MLE in practice, because **misspecification** may lead to poor estimates.

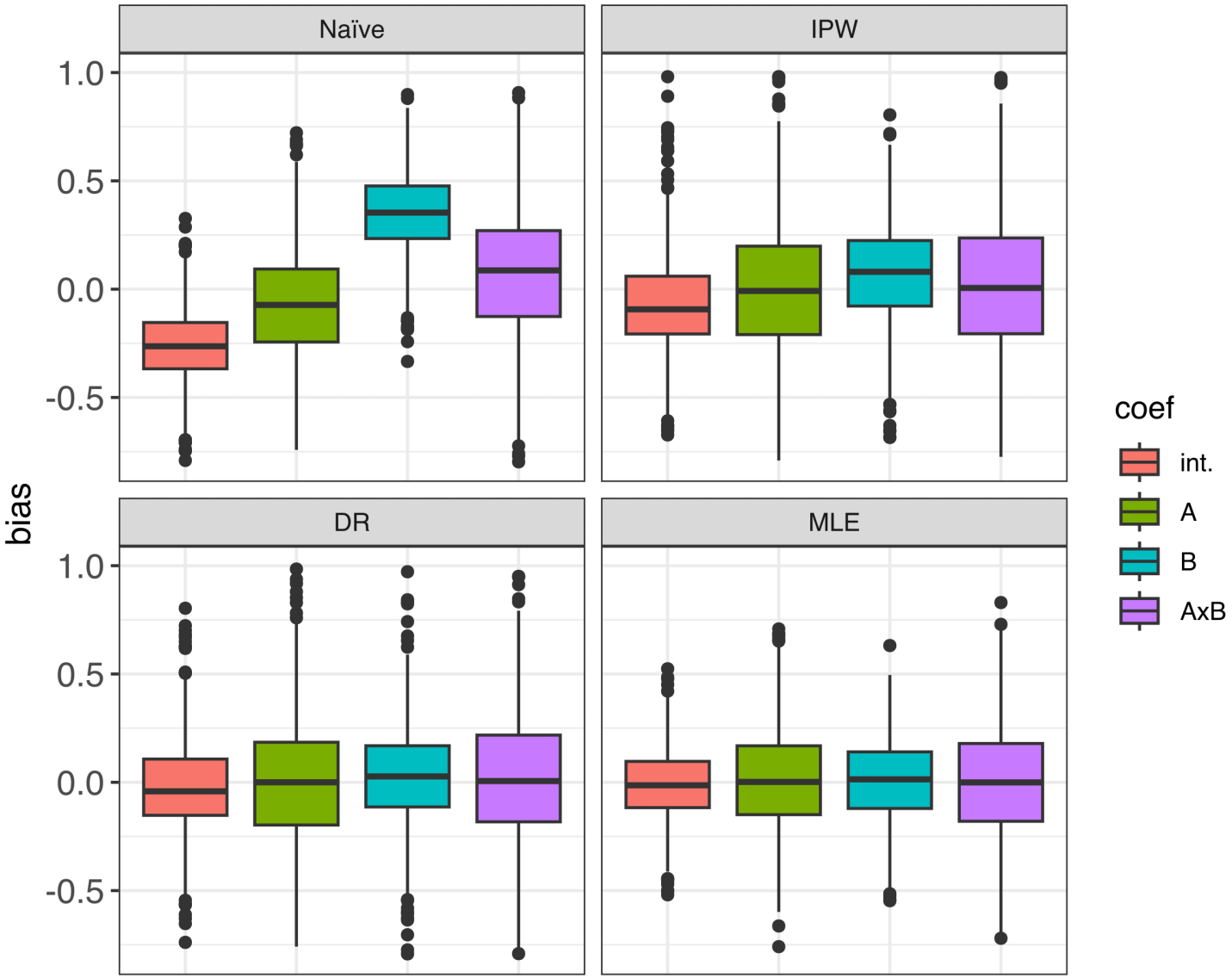
Results

Bias over 1,000 fits to simulated data ($n = 250$).



Results

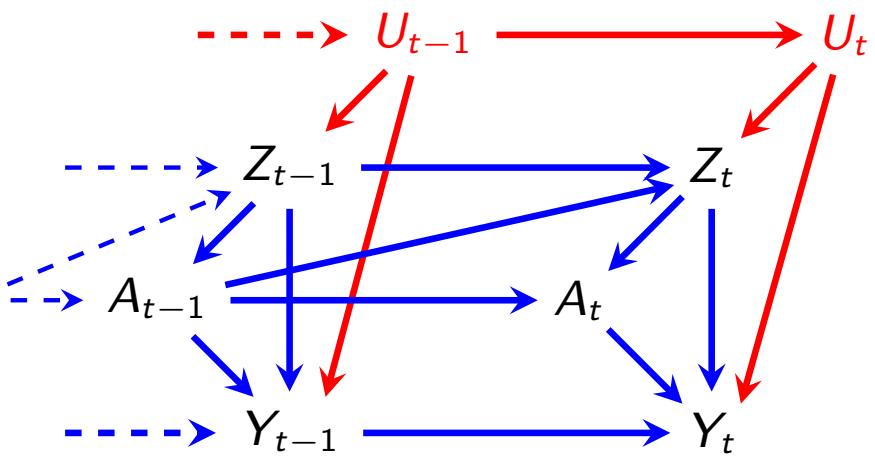
With a stronger $Z \to B$ edge ($\mathbb{E}[B \mid A, Z] = \text{expit}(a/2 + z)$):



Outline

Survival Analysis

We can also parameterize, simulate from, and fit survival models, such as **Cox Marginal Structural Models** with our method.



Specify a model for:

$$P(Y_t \mid Y_{t-1} = 0; do(A_1, \ldots, A_t))$$

where we marginalize over time-varying covariates $Z_1, \ldots, Z_t$.

This resolves an open problem in the literature: methods for simulating from Cox MSMs have been proposed, but they cannot **specify the marginal structure** in as much generality we can (e.g. Keogh et al., 2021).

Example: Survival Models

With the frugal parameterization simulation is easy even under a null hypothesis; e.g.:

$$P(Y_t \mid Y_{t-1} = 0; do(a_1, \ldots, a_t)) = P(Y_t \mid Y_{t-1} = 0).$$

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

$$P(Y_t \mid Y_{t-1} = 0; do(A_t = a)) = g(a).$$

Young and Tchetgen Tchetgen (2014) note that this is **not at all** trivial.

> *"We therefore may be limited to simulation scenarios with the proposed algorithm to unrealistic settings if we wish simultaneously to generate data under the null."*

Sensitivity Analysis

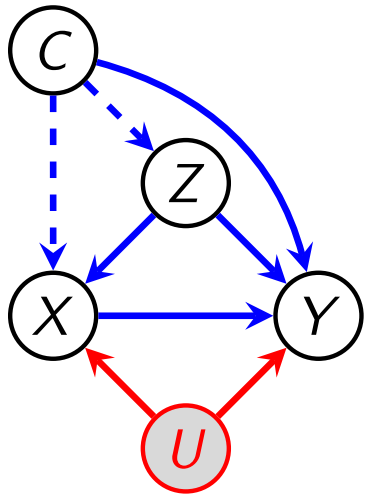
Note that since we have control over the propensity score model, it is comparatively easy to evaluate the effect of using the wrong model.

Example

Suppose we fit using a logistic regression model, but the true propensity uses a different link function (e.g. probit).

Or perhaps the truth is a random forest model.

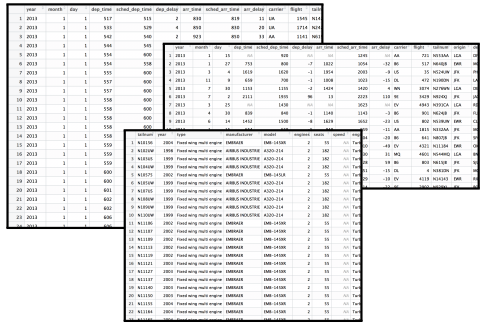
We can also see the effect of having unobserved confounders that are not included in the model.



'Many Data'

A more fundamental use for frugal parameterization is to allow for the **integrated causal analysis** of different types of studies.

Want to leverage data from **multiple sources** (different subjects, populations, and experimental settings) to improve causal learning.



- randomized controlled trials;
- observational cross-sectional and longitudinal studies;
- case-control studies;
- ...

We call this the paradigm of **Many Data**.

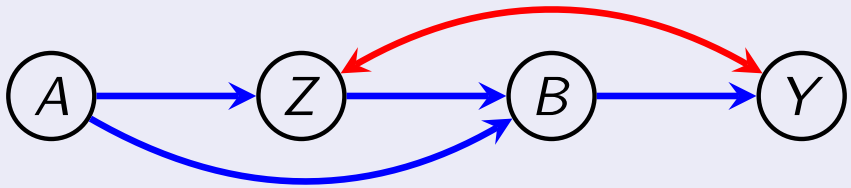
Using our parameterization we can **jointly** describe these models (e.g. Lin and Evans, 2023).

The Verma Constraint

Richardson et al. (2023) consider **nested** Markov models, which allow for **generalized** conditional independence constraints.

The frugal parameterization allows us to fit (some) nested models very easily.

Example



Constraint is that

$$p(y \mid do(a, b)) = \sum_z p(z \mid a) \cdot p(y \mid a, z, b)$$

is independent of A.

We can explicitly fit and test this constraint using (e.g.) likelihood ratio.

Summary

- We have presented the **frugal parameterization**, and used methods from marginal modelling to **simulate from** causal models;
- we can also **fit** these models using likelihood-based methods;
- this is a **marginal** parameterization: there is a rich literature on marginal models to consider for other causal problems.

- We envisage applications to marginal structural models, survival models, dynamic treatment regimes, structural nested models, stationarity, transportability, sensitivity analysis, data fusion (Many Data) ...;
- can also simulate from arbitrary instrumental variables models.

Limitations

- Mediation models are still difficult to simulate from!
- With continuous outcomes simulation (generally) relies on rejection sampling, which is inefficient in higher dimensions.

  In fact, can now use **inversion** of the conditional copula to simulate much more efficiently (linear in both dimension and sample size).

Thank you!

References I

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

Evans. `causl`, R package. `https://github.com/rje42/causl`, 2021.

Evans and Didelez. Parameterizing and Simulating from Causal Models (with discussion), *JRSS-B*, 2023.

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

Keogh, Seaman, Gran, Vansteelandt. Simulating longitudinal data from marginal structural models using the additive hazard model, *Biom. J.*, 2021.

Lin and Evans. Combine data from randomized trials

Richardson, Evans, Robins, Shpitser. Nested Markov properties for acyclic directed mixed graphs, *Ann. Statist.*, 2023.

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. Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, (pp. 95–133). 2000.

References II

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

Young, Hernán, Picciotto and Robins. Simulation from Structural Survival Models under Complex Time-Varying Data Structures. *JSM Proceedings, Section on Statistics and Epidemiology*, 2008.

Young, Hernán, Picciotto and Robins. Relation between three classes of structural models for the effect of a time-varying exposure on survival. *Lifetime Data Analysis*, 2010.

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

Joint distribution

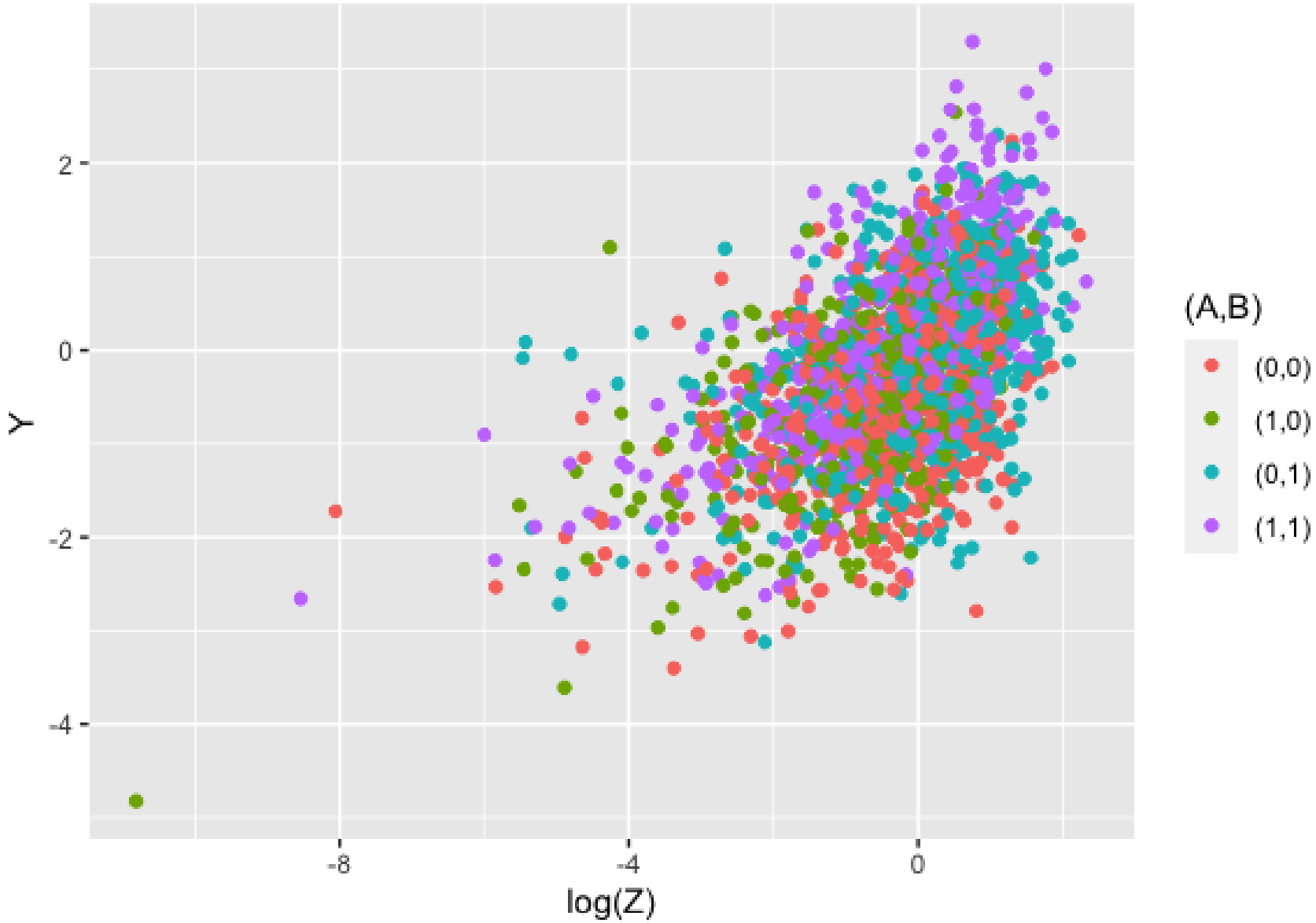
Let P be the distribution of interest, P^* its interventional counterpart.

We have

$$\begin{aligned} p^*(z,x,y) &= p(z,x,y) \cdot \frac{p^*(z,x)}{p(z,x)} \\ &= p(z,x,y) \cdot \frac{p^*(x) \cdot w(z \mid x)}{p(x,x)} \\ &= p^*(x) \cdot w(z \mid x) \cdot p(y \mid z,x). \end{aligned}$$

Note this factorization is not in the (standard) causal order.

Plot of Data



Structural Nested Mean Models

A **structural nested mean model** is defined by considering *blips* of treatment at each time-point; e.g.

$$\theta(\overline{z}_t, \overline{x}_{t-1}) := b_t(\overline{z}_t, \overline{x}_{t-1}, 1) - b_t(\overline{z}_t, \overline{x}_{t-1}, 0)$$

where

$$b_t(\overline{z}_t, \overline{x}_{t-1}, x) := \mathbb{E}[Y \mid \overline{z}_t, \overline{x}_{t-1}; do(X_t = x, \underline{X}_{t+1} = 0)].$$

These models are more flexible than marginal structural models, as they allow for the incorporation of the covariate history into the causal effect.

Structural Nested Mean Models

We can also parameterize this using a frugal parameterization at each time t.

Definition

Consider for $t = 1, \ldots, T$:

- $P(z_t, x_t \mid \overline{z}_{t-1}, \overline{x}_{t-1})$ (i.e. 'the past');
- $\theta(\overline{z}_t, \overline{x}_{t-1})$ (the parameter of interest);
- a conditional dependence measure between Y and Z_t given $\overline{X}_t, \overline{Z}_{t-1}$.

Then one can see that by building up from time $t-1$ to time t we go from

$$\mathbb{E}[Y \mid \overline{z}_{t-1}, \overline{x}_{t-1}; do(\underline{0}_t)] \qquad \text{to} \qquad \mathbb{E}[Y \mid \overline{z}_t, \overline{x}_t; do(\underline{0}_{t+1})];$$

i.e. the same thing with t replaced by $t+1$.

Example

Suppose we wish to model

$$Y \mid do(X = x) \sim \text{Gamma}(\mu_x, \phi\mu_x^2)$$

where $\mathbb{E}[Y \mid do(X = x)] = \mu_x = \exp(\beta_0 + \beta_1 x)$; along with specifying that

$$Z \sim \text{N}(\nu,\ \tau^2),$$
$$\log X \mid \{Z = z\} \sim \text{N}(\alpha_0 + \alpha_1 z, \sigma^2)$$

and that there is a Gaussian copula between Y and Z with partial correlation $2\,\text{expit}(\gamma_0 + \gamma_1 x) - 1$.

This specification is guaranteed to give a unique joint distribution, for any values of $\nu, \tau^2, \alpha_0, \alpha_1, \beta_0, \beta_1, \phi, \gamma_0, \gamma_1$ and σ^2.

Example

Suppose we pick:

$$\alpha_0 = -1 \quad \alpha_1 = 1 \qquad \beta_0 = -4 \quad \beta_1 = 0.5$$
$$\gamma_0 = 0.5 \quad \gamma_1 = 0.02 \qquad \nu = 0 \quad \sigma^2 = \tau^2 = 1 \qquad \phi = 2$$

We can simulate very quickly to obtain (say) 10^4 observations from P^*.

Copula Model Example

Recall our simulation with $n = 10^4$ observations.

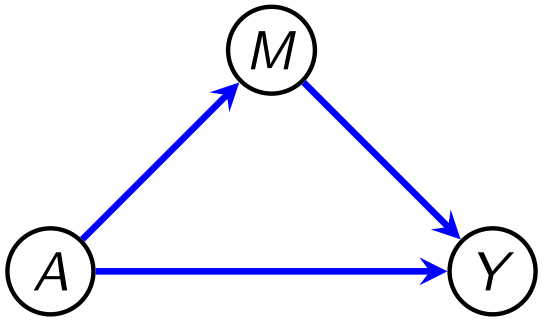
If we fit an ordinary (i.e. unweighted) linear model with

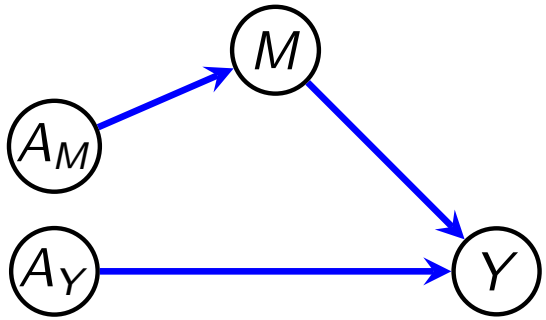
$$\mathbb{E}[Y \mid A = a, B = b] = \beta_0 + \beta_A a + \beta_B b + \beta_{AB} ab,$$

then the results are wrong:

parameter	truth	estimate	bias	s.e.	z-value	p-value
intercept	-1.0	-1.120	-0.120	0.022	-5.53	3.16×10^{-8}
A	0.5	0.550	0.050	0.033	1.52	0.13
B	0.5	0.656	0.156	0.028	5.51	3.67×10^{-8}
$A \cdot B$	0.0	-0.061	-0.061	0.041	-1.47	0.14

Mediation Models





Non-parametric mediation models typically ask what would happen if distinct treatment values were passed to the outcome Y and the mediator M, so they ask (e.g.) about the **natural** (in)direct effect (NDE/NIE):

$$NDE = \mathbb{E}[Y \mid do(A_M = 0, A_Y = 1)] - \mathbb{E}[Y \mid do(A_M = 0, A_Y = 0)]$$
$$NIE = \mathbb{E}[Y \mid do(A_M = 1, A_Y = 1)] - \mathbb{E}[Y \mid do(A_M = 0, A_Y = 1)].$$

The difficulty for the frugal parameterization is that we must enforce $Y \perp\!\!\!\perp A_M \mid A_Y, M$ **and** model $p(y \mid a_M, a_Y)$; this does not usually lead to congenial models.

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 to note:

- 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).

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

Odds Ratio Examples

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

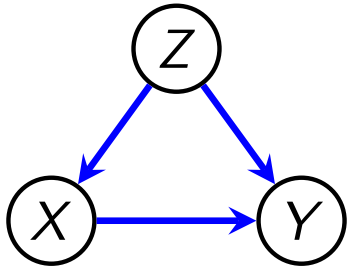
$$\phi_{XY} \sim \exp\left(\frac{\rho xy}{\sigma_x \sigma_y (1-\rho^2)}\right)$$

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

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

Margins

Let's think about the simplest example of this kind.



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

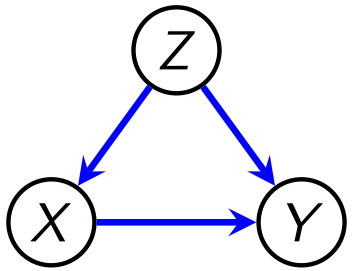
This is a 'margin' of the joint distribution

$$p^*(z, y \mid x) \equiv p(z)p(y \mid x, z).$$

To work with P^* we need to model the XY-margin (because that's the quantity of interest) and the XZ-margin (to enforce the independence).

So what's left to know?

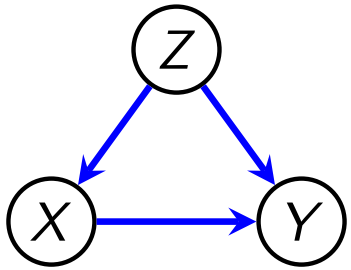
Odds Ratios



Bergsma and Rudas' results show that the remaining information is precisely the odds ratio between Y and Z conditional upon X.

Any additional information given the dependence ratio, $p(y \mid do(x))$, and $p(x, z)$ would be redundant.

Odds Ratios



There's nothing to stop us specifying that the parameters β and γ are from this model:

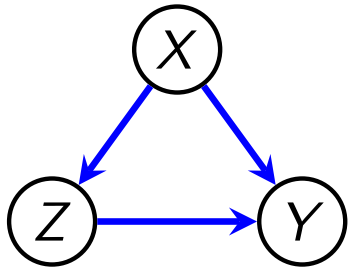
$$\text{logit}\, p(y \mid x, z) = \mu + \alpha x + \beta z + \gamma xz.$$

But μ and α are **not free**.

Take home message—you can have part of a nice model on X, Y, Z that includes $p(y \mid do(x))$; just don't expect all of it!

g-null Paradox Illustration

Suppose that we have continuous X and Y, but binary Z.



An innocuous seeming model would be:

$$\mathbb{E}[Y \mid X = x, Z = z] = \mu + \beta x + \gamma z.$$

But:

$$\begin{aligned}\mathbb{E}[Y \mid X = x] &= \sum_z \mathbb{E}[Y \mid X = x, Z = z] \cdot P(Z = z \mid X = x) \\ &= \mu + \beta x + \gamma P(Z = 1 \mid X = x).\end{aligned}$$

Now $P(Z = 1 \mid X = x)$ can't be a linear function of x (unless it's constant). So $\mathbb{E}[Y \mid X = x]$ is only a linear function if either:

- $Z \perp\!\!\!\perp X$; or
- $\gamma = 0$ (so $Y \perp\!\!\!\perp Z \mid X$).