Parameterizing and Simulating from Causal Models

by Robin Evans and Vanessa Didelez

Zhiling Gu

Iowa State University

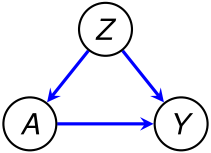
May 7, 2023

This discussion of Evans and Didelez (2021) is based on a Online Causal Conference Seminar, this Recording and this Recording.

Outline

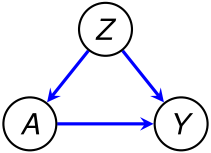
Example

Consider X: diet, Y: BMI, Z: indicator of education level.



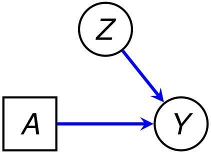
Example

Consider X: diet, Y: BMI, Z: indicator of education level.



If we performed an experiment where $A = a$ is set by external intervention,

- the randomized A removes confounding effect from Z, i.e. $A \perp\!\!\!\perp Z$.
- marginal dist $[Z]$, conditional dist $[Y|A, Z]$ are preserved.



The Problem

As in Marginal Structural Models (MSM) by Robins et al. (2000), we consider the potential outcome of Y given $X = x$, i.e., the marginal effect of X on Y. Under SUTVA, SRA, and Positivity:

$$P^*(Y = y|A = a) = \sum_z P(Z = z)P(Y = y|Z = z, A = a), \tag{1}$$

which is also denoted as $P(Y = y|do(A = a))$.

- $[\cdot]^*$: distribution/parameter from causal or interventional distribution
- $[\cdot]$: distribution/parameter from observational regime

Example R1

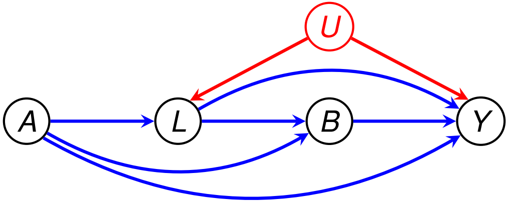


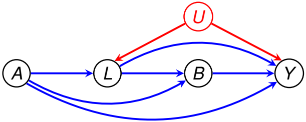
Figure 1: The causal model from Havercroft and Didelez (2012).

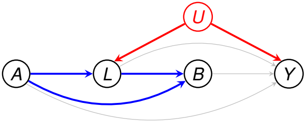
- The second treatment B depends on both the first treatment A and an intermediate outcome L
- The variable U is 'hidden' or latent
- Identifiable quantities are functions of P_{ALBY}
- Quantity of interest: conditional distribution $P_{Y|A,B}$

g-formula (Robins, 1986): Under the assumption of positivity and the causal structure implied by the graph ($p_{L|AB} = p_{L|A}$), $p_{Y|AB}$ is identified by the g-formula

$$p_{Y|AB}(y \mid do(a,b)) := \int p_{Y|ALB}(y \mid a, \ell, b) \cdot p_{L|A}(\ell \mid a)\, d\ell. \tag{2}$$

g-null: $H_0 : A, B$ has no effect on Y, i.e. there are no arrows from A to Y, nor B to Y.





g-null Paradox

g-null Paradox (Robins and Wasserman, 2013): $p_{Y|AB}$ would depend on A even if g-null is true, i.e. A, B has no causal effect on Y.
Example: Suppose

$$Z|A \sim Ber(expit(\alpha A)),\ \mathbb{E}[Y|A, Z, B] = A\beta_a + Z\beta_z + B\beta_b,$$

then

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

For $\mathbb{E}[Y|do(A, B)]$ to be independent from A, we need

$$\beta_a = 0 \text{ and } \alpha\beta_z = 0,$$

which is equivalent to

$$\text{either } Y \perp\!\!\!\perp A, Z|B \text{ or } \begin{Bmatrix} Y \perp\!\!\!\perp A|B \\ A \perp\!\!\!\perp Z \end{Bmatrix}$$

Models avoiding g-null Paradox

J.Robins invented many causal models such as marginal structural models (MSM) and structural nested models (SNM) to make it easier to estimate causal effects. These models

- they lead to tractable estimators
- they are semiparametric
- they avoid the g-null paradox

Marginal Structural Models (MSM)

MSM was proposed to work on longitudinal data, where there are time-dependent confounders, and treatments depend on historical data. Given a time series of covariates Z_t, treatments A_t, and responses Y_t:

$$\{(Z_t, A_t, Y_t)\}_{t=1}^{T}.$$

An MSM:

$$g(a_1, \ldots, a_T) = \mathbb{E}(Y(a_1, \ldots, a_T)) = \beta_0 + \beta_1 \sum_t a_t,$$

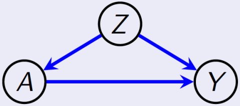
where β_0, β_1 can be easily estimated without invoking the g-null paradox.

Setup

In general, consider the following setting

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

'cartoon'

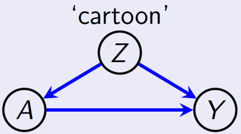


Objective: find $p_{Y|do(A)} \equiv P^*(Y|A)$.

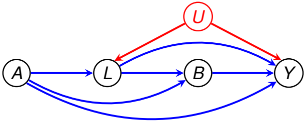
Setup

In general, consider the following setting

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



Objective: find $p_{Y|do(A)} \equiv P^*(Y|A)$.
There is no strict causal order on A, Z, Y. In Example R1, $A' = (A, B), Z' = L, Y' = Y$. Confounder U is unobserved.



Cognate Probabilities

We say $P^*(y|a)$ **is cognate to** $P(y|a)$(within $P(z,a,y)$) if $\exists w(z|a)$ s.t.,

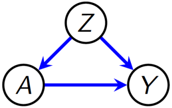
$$P^*(y|a) = \int \underbrace{P(y|a,z)w(z|a)}_{P^*(y|a,z)} dz,$$

where $w(z|a)$ is a kernel function such that

$$w(z|a) \geq 0, \quad \int w(z|a)dz = 1, \forall a.$$

Cognate Probabilities: Examples

Y: diabetes, A: treatment, Z: genetic information



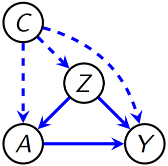
$$P(y|a) = \int P(y|a,z)P(z|a)dz$$

$$P^*(y|a) = P(y|do(a)) = \int P(y|a,z)P(z)dz$$

Thus, $P^*(y|a)$ is cognate to $P(y|a)$ by treating $P(z)$ as kernel function.

Cognate Probabilities: Examples

Y: diabetes, A: treatment, C: gene, Z: associated economics data.



$$P_c^*(y|a) = P(y|c, do(a)) = \int P(y|a, z, c)P(z|c)dz$$

Therefore, $P_c^*(y|a)$ is cognate to $P(y|a, c)$ by treating $P(z|c)$ as the kernel function.

Cognate Probabilities: Examples

We can calculate the probability of **potential outcome** of Y when A is set to a, given observed $A = a'$,

$$P^*(Y(a)|a') = \int P_{Y|ZA}(y|z,a)P_{Z|A}(z|a')dz$$

where $P^*(Y(a)|a')$ is cognate to $P(Y(a)|a')$.

Effect of Treatment on the Treated (ETT):

ETT $= \mathbb{E}[Y(1) \mid A = 1] - \mathbb{E}[Y(0) \mid A = 1]$, where

$$\begin{aligned}\mathbb{E}(Y(a)|a') &= \int\int yP_{Y|ZA}(y|z,a)P_{Z|A}(z|a')\,dy\,dz \\ &= \int y \underbrace{\int P_{Y|ZA}(y|z,a)P_{Z|A}(z|a')\,dz}_{\text{Cognate to } P_{Y|ZA}}\,dy\end{aligned}$$

Reformulate Cognate Probabilities

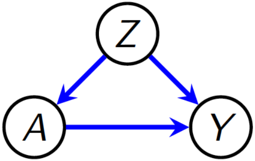
$$
\begin{aligned}
P^*_{Y|A}(y|a) &= \int P_{Y|ZA}(y|z,a)w(z|a)dz \\
&= \int \frac{P_{ZAY}(z,a,y)}{P_{ZA}(z,a)} w(z|a)dz \\
&= \int \frac{P^*_{ZAY}(z,a,y)}{P^*_{ZA}(z,a)} w(z|a)dz,
\end{aligned}
$$

where

$$
\begin{aligned}
P^*_{ZAY}(z,a,y) &= P_{ZAY}(z,a,y)\frac{P^*_{ZA}(z,a)}{P_{ZA}(z,a)} \\
&= P_{ZAY}(z,a,y)\frac{P^*_{A}(a)w(z|a)}{P_{ZA}(z,a)} \\
&= P_{Y|ZA}(y|z,a)P^*_{A}(a)w(z|a)
\end{aligned}
$$

Therefore $P^*_{Y|A}$ can be seen as taken from P^*_{ZAY}.

Frugal Parameterization



To identify observational distribution p_{XYZ}, decompose it into 3 parts

- P_{ZA}: the *past*
- $P^*_{Y|A}$: the *causal* distribution of interest
- $\phi^*_{YZ|A}$: a *dependence measure*

Variation Independence

Given ϕ and ψ are two functions defined on Θ. We say ϕ and ψ are **Variation Independent** if

$$(\phi \times \psi)(\Theta) = \phi(\Theta) \times \psi(\Theta)$$

Variation Independence

Given ϕ and ψ are two functions defined on Θ. We say ϕ and ψ are **Variation Independent** if

$$(\phi \times \psi)(\Theta) = \phi(\Theta) \times \psi(\Theta)$$

(A1) Given a frugal parameterization

$$\theta = (\underbrace{\theta_{ZA}}_{\textit{past}}, \underbrace{\theta_{Y|A}}_{\textit{causal}}, \underbrace{\phi_{YZ|A}}_{\textit{association}}),$$

the parameter $\phi_{YZ|A}$ is jointly variation indenpendent of θ_{ZA} and $\theta_{Y|A}$.

- Ensures the space of joint distribution can be separated
- Not necessary for main result, but makes interpretation easier

Association parameterizations satisfying (A1)

If A, Y, Z are finite categorical variables,

- conditional odds ratio: $\phi_{YZ|A} := \frac{p(1,1|a)p(0,0|a)}{p(1,0|a)p(0,1|a)}$.

If A, Y, Z are multivariate Gaussian random variables/ distributions defined by first two moments,

- partial correlation: $\rho_{YZ|A} := Cor(Y, Z|A)$

If A, Y, Z are general continuous variables,

- conditional copula:
  $C_{YZ|A}(u, v|a) := \Pr(F_Y(Y) \leq u, F_Z(Z) \leq v|A = a), u, v \in [0, 1]$.

If A, Y, Z involve continuous and binary variables,

- use Gaussian conditional copula that is dichotomized for binary components Fan et al. (2017)

Remark: $\phi_{YZ|A}$ is parametric to specify the family of copulas of interest.

Main Result I

Given a parameterization of P_{ZAY}

$$\theta = (\underbrace{\theta_{ZA}}_{P_{ZA}}, \underbrace{\theta_{Y|A}}_{P_{Y|A}}, \underbrace{\phi_{YZ|A}}_{P_{YZ|A}}),$$

we can choose any parametric model for any **cognate** distribution $P^*_{Y|A}$ to construct a smooth frugal parameterization of P_{ZAY},

$$\theta^* = (\underbrace{\theta_{ZA}}_{P_{ZA}}, \underbrace{\theta^*_{Y|A}}_{P^*_{Y|A}}, \underbrace{\phi^*_{YZ|A}}_{P^*_{YZ|A}}).$$

- $P^*_{Y|A}$ is determined by the analyst based on subject matter
- θ and θ^* are not generally the same
- θ and θ^* correspond to different interpretations

Main Result II

(A2) The product $P^*_{ZA} = w \cdot P^*_A$ has a smooth and regular parameterization $\eta_{ZA} := \eta_{ZA}(\theta_{ZA})$, where η_{ZA} is a twice differentiable function with a Jacobian of constant rank.

(A3) (Positivity) P_{ZA} is absolutely continuous w.r.t. P^*_{ZA} at true distribution P_{ZAY},

$$P^*_{ZA}(z,a) = 0 \implies P_{ZA}(z,a) = 0, \; \forall z, a$$

Main Result II

(A2) The product $P^*_{ZA} = w \cdot P^*_A$ has a smooth and regular parameterization $\eta_{ZA} := \eta_{ZA}(\theta_{ZA})$, where η_{ZA} is a twice differentiable function with a Jacobian of constant rank.
(A3) (Positivity) P_{ZA} is absolutely continuous w.r.t. P^*_{ZA} at true distribution P_{ZAY},

$$P^*_{ZA}(z,a) = 0 \implies P_{ZA}(z,a) = 0, \ \forall z, a$$

Theorem 3.1: Under (A2) – (A3), We can smoothly parameterize the joint distribution P with a frugal parameterization of

$$P_{ZA}, \ P^*_{Y|A}, \ \phi^*_{YZ|A}$$

if and only if P can be smoothly parameterized by the same models applied to

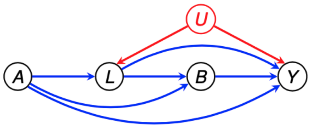
$$P_{ZA}, \ P_{Y|A}, \ \phi_{YZ|A}.$$

One-line proof: $P^*_{ZAY} = P_{ZAY} \frac{P^*_{ZA}}{P_{ZA}} = P_{ZAY} \frac{w \cdot P^*_A}{P_{ZA}}$

Example R1: Parameterizing

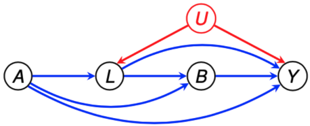
Take $Z = L$, $X = (A, B)$, then we can parameterize P_{ALBY} using

$$P_{ALB}(a, \ell, b) \quad P^*_{Y|AB}(y|a, b) \quad \phi^*_{LY|AB}(\ell, y|a, b)$$



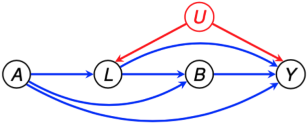
- $A \sim Ber(\theta_a)$, $L|A = a \sim Exp(\exp(-(\alpha_0 + \alpha_a a)))$
- $B|A = a, L = \ell \sim Ber(\text{expit}(\gamma_0 + \gamma_a a + \gamma_\ell \ell + \gamma_{a\ell} a\ell))$
- $LY|do(A = a, B = b) \sim BiGaussian$ with correlation parameter ρ_{ab}
- $Y|do(A = a, B = b) \sim N(\beta_0 + \beta_a a + \beta_b b + \beta_{ab} ab, \sigma^2)$

Example R1: Sampling



[Step 1] $A \sim Ber(\theta_a)$, $B \sim Ber(\theta_b)$
[Step 2] Generate correlated quantiles for L, Y from copula $LY|do(A=a, B=b) \sim BiGaussian$ with correlation parameter ρ_{ab}.
[Step 3] Generate Y and L using Inverse CDF (using quantiles from Step 2) from
$Y|do(A=a, B=b) \sim N(\beta_0 + \beta_a a + \beta_b b + \beta_{ab} ab, \sigma^2)$,
$L|A=a \sim Exp(\exp(-(\alpha_0 + \alpha_a a)))$
$\implies$ Intervened Distribution of Y

Example R1: Sampling



[Step 1] $A \sim Ber(\theta_a)$, $B \sim Ber(\theta_b)$
[Step 2] Generate correlated quantiles for L, Y from copula $LY|do(A = a, B = b) \sim BiGaussian$ with correlation parameter ρ_{ab}.
[Step 3] Generate Y and L using Inverse CDF (using quantiles from Step 2) from
$Y|do(A = a, B = b) \sim N(\beta_0 + \beta_a a + \beta_b b + \beta_{ab} ab, \sigma^2)$,
$L|A = a \sim Exp(\exp(-(\alpha_0 + \alpha_a a)))$
$\implies$ Intervened Distribution of Y
[Step 4] Generate B using rejection sampling such that
$B|A = a, L = \ell \sim Ber(\mathrm{expit}(\gamma_0 + \gamma_a a + \gamma_\ell \ell + \gamma_{a\ell} a\ell))$.
$\implies$ Observational Distribution of Y

Maximum Likelihood Estimation (MLE)

MLE for $P^*_{Y|A}$ is obtained by maximizing the likelihood for the causal model w.r.t observational joint distribution P_{ZAY}.

Maximum Likelihood Estimation (MLE)

MLE for $P^*_{Y|A}$ is obtained by maximizing the likelihood for the causal model w.r.t observational joint distribution P_{ZAY}.

(A5) $\mathrm{KL}(P_{ZA}||P^*_{ZA}) := \mathbb{E}_{P_{ZA}}\{\log \frac{P_{ZA}(z,a)}{P^*_{ZA}(z,a)}\} < \infty$.

Theorem 5.1 Suppose θ^* is a frugal parameterization with weight function $w(z) = p_Z(z)$ (i.e. the MSM model); and (A5) holds. Then MLE $\hat{\eta}$ of $\eta(\theta^*)$ obtained with the observed data (i.e. data generated using distribtion P_{ZAY} with parameters $\theta^* = (\theta_{ZA}, \theta^*_{Y|A}, \phi^*_{YZ|A})$) will be consistent for the distribution in the causal model with parameters $\eta = (\eta_{ZA}(\theta_{ZA}), \theta^*_{Y|A}, \phi^*_{YZ|A})$,

$$\sqrt{n}\left\{\begin{pmatrix}\hat{\theta}^*_{Y|A}\\ \hat{\phi}^*_{YZ|A}\end{pmatrix} - \begin{pmatrix}\theta^*_{Y|A}\\ \phi^*_{YZ|A}\end{pmatrix}\right\} \implies N(0, I(\theta^*)^{-1}_{\theta^*_{Y|A},\phi^*_{YZ|A}})$$

Maximum Likelihood Estimation (MLE)

MLE for $P^*_{Y|A}$ is obtained by maximizing the likelihood for the causal model w.r.t observational joint distribution P_{ZAY}.

(A5) $\mathrm{KL}(P_{ZA}||P^*_{ZA}) := \mathbb{E}_{P_{ZA}}\{\log \frac{P_{ZA}(z,a)}{P^*_{ZA}(z,a)}\} < \infty$.

Theorem 5.1 Suppose θ^* is a frugal parameterization with weight function $w(z) = p_Z(z)$ (i.e. the MSM model); and (A5) holds. Then MLE $\hat{\eta}$ of $\eta(\theta^*)$ obtained with the observed data (i.e. data generated using distribtion P_{ZAY} with parameters $\theta^* = (\theta_{ZA}, \theta^*_{Y|A}, \phi^*_{YZ|A})$) will be consistent for the distribution in the causal model with parameters $\eta = (\eta_{ZA}(\theta_{ZA}), \theta^*_{Y|A}, \phi^*_{YZ|A})$,

$$\sqrt{n}\left\{\begin{pmatrix}\hat{\theta}^*_{Y|A}\\ \hat{\phi}^*_{YZ|A}\end{pmatrix} - \begin{pmatrix}\theta^*_{Y|A}\\ \phi^*_{YZ|A}\end{pmatrix}\right\} \Longrightarrow N(0, I(\theta^*)^{-1}_{\theta^*_{Y|A}, \phi^*_{YZ|A}})$$

Remark: When the models are correctly specified, MLE is the most efficient. Otherwise, IPW and AIPW(Doubly Robust) are recommended in practice.

Estimators

- Naïve:

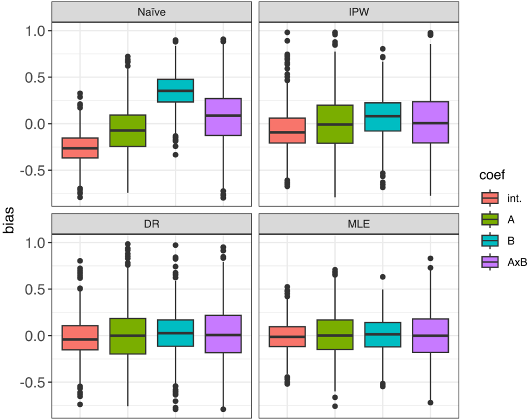
$$\mathbb{E}[Y \mid A = a, B = b] = \beta_0 + \beta_A a + \beta_B b + \beta_{AB} ab$$

- IPW:

$$\text{logit} P(B = 1 | A = a, L = \ell) = \alpha_0 + \alpha_A a + \alpha_L \ell,$$

$$\widehat{Y}_{\text{IPW}}(A = a, B = b) = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i(a_i = a, b_i = b) Y_i}{\hat{P}_{B|AL}(b|a, \ell_i))}$$

Simulation Results



- Naïve outcome regressor does not work
- IPW, DR, MLE work well when the model is correctly specified
- MLE is more efficient when model correctly specified

Discussion

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^*.)

The frugal parameterization resolves these as best one can.

Thank you!

References I

Robin J Evans and Vanessa Didelez. Parameterizing and simulating from causal models. arXiv preprint arXiv:2109.03694, 2021.

James M Robins, Miguel Angel Hernan, and Babette Brumback. Marginal structural models and causal inference in epidemiology. Epidemiology, pages 550–560, 2000.

WG Havercroft and Vanessa Didelez. Simulating from marginal structural models with time-dependent confounding. Statistics in medicine, 31(30):4190–4206, 2012.

James Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. Mathematical modelling, 7(9-12): 1393–1512, 1986.

James M Robins and Larry A Wasserman. Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs. arXiv preprint arXiv:1302.1566, 2013.

References II

Jianqing Fan, Han Liu, Yang Ning, and Hui Zou. High dimensional semiparametric latent graphical model for mixed data. Journal of the Royal Statistical Society. Series B (Statistical Methodology), pages 405–421, 2017.

Takeaway and Discussion I

Traditional problem of interest: Given a joint distributioni P, we can apply the g-formula to get the causal distribution P^* for the counterfactual Y^*,

$$P \xrightarrow{g} P^*$$

The reverse problem

$$P^* \xrightarrow{?} P.$$

g-formula

$$P^*(y|a) = \int P(y|x, a) dP(x)$$

where

$$P_a(y) = P^*(y|a) = P(y|do(A = a))$$

Another Perspective of he Frugal Construction

$$p(x, a, y) = p(x, a)p(y|x, a) \overset{do(a)}{=} p(x, a)p_a(y|x) = p(x, a)\frac{p_a(x, y)}{p_a(x)}$$

$$= p(x, a)\frac{p_a(x)p_a(y)c_a(x, y)}{p_a(x)} = p(x, a)p_a(y)c_a(x, y).$$

where $p_a(y)$ can be parameterized using MSM, $p_a(y; \beta)$. This construction avoids the specification of both $p_a(x), p_a(y)$.